



03061

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

U. A. C. P. Y P. del C.C.H.

29

Bondad de ajuste en modelos logísticos

T E S I S

QUE PARA OBTENER EL TITULO DE

MAESTRO EN CIENCIAS

Maestro en estadística e investigación de operaciones.

PRESENTA

ALEJANDRO ALEGRIA HERNANDEZ

MEXICO, D.F. 1995

FALLA DE ORIGEN



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

INTRODUCCION	1
1. EL PROBLEMA DE EVALUACION EN MODELAJE	3
2. MODELO LOGISTICO	6
2.1 DEFINICION	6
2.2 MODELOS LINEALES GENERALIZADOS	9
3. ESTADISTICA TIPO DIVERGENCIA	17
3.1 DATOS AGRUPADOS Y NO AGRUPADOS	17
3.2 ESTADISTICA TIPO DIVERGENCIA	19
3.3 RESULTADOS ASINTOTICOS	21
3.3.1 RESULTADOS m -ASINTOTICOS	22
3.3.2 RESULTADOS n -ASINTOTICOS	28
3.3.2.1 DISPERSION EXTREMA	34
4. APLICACION DE RESULTADOS	38
5. CONCLUSIONES	46
APENDICE 1	49
APENDICE 2	51
APENDICE 3	53
APENDICE 4	54
BIBLIOGRAFIA	56

INTRODUCCION

Siempre ha sido interesante el problema de explicar, aunque sea de forma parcial, la realidad que nos rodea. Los modelos estadísticos son una importante herramienta que nos permiten expresar, en términos simbólicos, posibles relaciones entre variables. Una vez que se ha propuesto un modelo, se requiere de la información proporcionada por un conjunto de observaciones, con el propósito de ajustar dicho modelo, y hacer uso de él posteriormente. ¿Cómo saber qué tan bueno es un ajuste? Una forma de responder a esta pregunta es utilizando lo que en Estadística se conoce como pruebas de Bondad de Ajuste. Este tipo de pruebas tratan de ser más objetivas que un simple análisis gráfico, y por lo general requieren del conocimiento de cierta herramienta matemática. El evaluar qué tan bueno resulta un modelo es sólo una parte del proceso de construcción de modelos. En el Capítulo 1 de este trabajo se comenta brevemente en qué consiste este proceso.

Existe mucha literatura sobre el problema de Bondad de Ajuste cuando se estudian variables en forma individual, pero la situación es más complicada cuando los modelos propuestos involucran a más de una variable. En este trabajo el objetivo es presentar, en forma no tan rigurosa, algunos resultados que se cree pueden aportar cierta claridad en cuanto al uso de algunas estadísticas de prueba ya conocidas. Los modelos que más nos interesan son los llamados logísticos, y por esta razón en el Capítulo 2 se define el modelo logístico, se mencionan algunas de sus propiedades y se hace ver que este modelo pertenece a la familia de los Modelos Lineales Generalizados. En estos términos se obtienen las expresiones adecuadas de los estimadores, mismas que aplicarán posteriormente.

En el Capítulo 3 se presenta la estadística que nos permitirá realizar las pruebas de bondad de ajuste sobre el modelo logístico. Es importante mencionar que, en principio, el modelo puede ser cualquiera de la familia de los modelos lineales generalizados. La distinción entre tener datos agrupados y no agrupados es crucial, pues la estadística de prueba tiene una distribución diferente en cada caso.

La forma de la estadística de prueba que se va a utilizar está basada en el concepto de divergencia, y adicionalmente usa potencias en las comparaciones que realiza. Esto hace muy difícil el encontrar su distribución exacta, y por esta razón los resultados que se presentan en dos teoremas, son de tipo asintótico.

Dos ejemplos conforman el Capítulo 4, uno utiliza datos agrupados y el otro datos no agrupados.

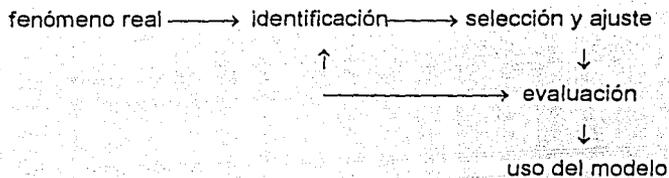
Finalmente se dan algunas conclusiones, en donde se plantean algunas ventajas y desventajas del uso de la estadística presentada.

En los Anexos se pueden encontrar algunas demostraciones y también las condiciones que teóricamente deben cumplirse para que sea válida la afirmación que se hace en los Teoremas.

CAPÍTULO 1

EL PROBLEMA DE EVALUACION EN MODELAJE

El proceso de construcción de un modelo estadístico se puede esquematizar como sigue,



En la etapa de identificación se busca definir una familia de modelos a partir de un análisis exploratorio de la información y de un manejo adecuado de los elementos fundamentales del fenómeno que se pretende modelar. La segunda etapa consiste en seleccionar el mejor modelo dentro de la familia que ya se tiene; es aquí donde se aplican técnicas de estimación. En la siguiente etapa se adopta una postura crítica con respecto a qué tan adecuadamente el modelo ajustado describe el fenómeno bajo estudio. Esto último se logra realizando una evaluación del modelo a partir de la cual no sólo se puede llegar a decir que éste resulta inadecuado, sino también el porqué de tal inadecuación. Si los métodos de evaluación indican que el modelo no es capaz de explicar significativamente el fenómeno de interés, se tendrán que hacer modificaciones en las siguientes etapas del proceso iterativo antes de que el modelo pueda ser usado.

Las técnicas de diagnóstico y las pruebas de bondad de ajuste son las herramientas estadísticas que nos permiten llevar a cabo el proceso de verificación de un modelo.

Con respecto al problema de la construcción de modelos McCullagh y Nelder (1989) dicen lo siguiente:

"Modelling in science remains, partly at last, an art. Some principles do exist, however to guide the modeller. A first, thought at first sight, not a very helpful principle, is that all models are wrong; some, though, are most useful than others and we should seek those. At the same time we must recognize that eternal truth is not within our grasp. A second principle (which applies also to artist) is not to fall in love with one model to the exclusion of alternatives. Data will often point with almost equal emphasis at several possible models and it is important that the statistician recognize and accept this. A third principle recommends through checks on the fit of a model to the data, for example by using residuals and other statistics derived from the fit, to look for outlying observations and so on. Such diagnostic procedures are not yet fully formalized, and perhaps never will be. Some imagination is required here in order to determine the aspects of the model that are more important and most suspect.."

En este trabajo nos interesa analizar lo que se ha hecho y lo que se puede hacer en la etapa de verificación de un modelo. Específicamente, nos avocaremos al problema de evaluar el ajuste de un modelo dado, entendiéndose con esto el poder realizar una contrastación entre la información que se tenga (generalmente datos) y el modelo propuesto.

Los métodos estadísticos desarrollados para evaluar el ajuste de un modelo se pueden clasificar como sigue: 1) los que pretenden cuantificar globalmente el ajuste y, 2) los que buscan aspectos específicos para mejorar el ajuste. También se puede hacer otra clasificación de los métodos estadísticos en cuanto a su grado de formalidad. Las técnicas que se basan en la construcción y

análisis de gráficas se consideran informales, no obstante constituyen un elemento importante en el proceso de evaluación de un ajuste. Las técnicas consideradas formales se desarrollan a partir de estimaciones que nos permiten probar si el modelo propuesto se ajusta bien a los datos o no, ya sea en términos globales (omnibus test) ó en aspectos más específicos (direccional test).

En la práctica, una vez que se han estimado los parámetros involucrados en un modelo, la etapa de evaluación se inicia con el cálculo de las medidas globales de ajuste. Si éstas indican que, en términos generales, el modelo es bueno, se procede a aplicar técnicas de diagnóstico, cuyos objetivos son, entre otros, identificar datos atípicos que pudieran o no ser influyentes, y determinar, en el caso de tener una variable respuesta, si la forma funcional que relaciona ésta última con las variables explicativas es la correcta. Es usual que las técnicas de diagnóstico indiquen qué correcciones deben hacerse, con lo cual es más factible que se tenga un modelo que explique mejor el fenómeno bajo estudio y/o se hagan predicciones más certeras sobre el comportamiento del mismo. (Varios son los trabajos que se han hecho sobre diagnóstico, principalmente en modelos de regresión: Chatterjee, Samprit y Hadi (1988); Cook, D.R. (1986); Cook, D.R. y Weisberg, S. (1982), Daniel y Wood (1971), entre otros.)

Como se dijo anteriormente, el presente trabajo pretende analizar lo que se ha hecho y lo que se puede hacer en la etapa de evaluación de un modelo, dando mayor énfasis al problema de construcción, propiedades y uso de medidas globales de ajuste. En particular, se estudiará el caso del modelo logístico, el cual puede ser visto como un caso particular de los llamados modelos lineales generalizados. En el próximo capítulo se definirá el modelo logístico y se mencionarán algunas de sus características.

CAPITULO 2

MODELO LOGISTICO

2.1 DEFINICION

El modelo que nos interesa analizar en este trabajo es conocido con el nombre de modelo logístico. Este modelo resulta de mucha utilidad al estudiar el comportamiento de una variable de respuesta dicotómica con respecto a los diferentes valores de algunas variables explicativas. Cuando estas últimas variables están medidas en una escala continua, lo que se tiene es un modelo de regresión logístico.

Para definir formalmente el modelo logístico, consideremos a Y , una variable aleatoria dicotómica tal que $P(Y=1) = \pi$ y W_1, W_2, \dots, W_s un conjunto de variables explicativas. Como una primera aproximación, podría sugerirse el modelo lineal

$$\pi = E(Y) = \sum_{j=0}^{p-1} \delta_j g_j(W_1, W_2, \dots, W_s), \quad \rightarrow$$

en donde el valor esperado está condicionado a los valores de W_1, W_2, \dots, W_s , las funciones g_j son conocidas con $g_0(W_1, W_2, \dots, W_s) \equiv 1$; y $\delta_0, \delta_1, \dots, \delta_{p-1}$, es un conjunto de parámetros, usualmente desconocidos. Si la información muestral consiste de n observaciones independientes, el modelo anterior se puede expresar como

$$\pi = E(Y) = X \delta,$$

donde

$$\pi^T = (\pi_1, \pi_2, \dots, \pi_n), \quad Y^T = (Y_1, Y_2, \dots, Y_n), \quad \delta^T = (\delta_0, \delta_1, \dots, \delta_{p-1}),$$

y X es la matriz de diseño de dimensión $n \times p$, cuyas columnas corresponden a los valores observados de las variables

$$x_j = g_j(W_1, W_2, \dots, W_s), \quad j = 0, 1, 2, \dots, p-1.$$

El modelo anterior se caracteriza principalmente por: a) la variable respuesta es dicotómica, b) el valor de δ tiene que restringirse pues π_i debe de estar en el intervalo $(0,1)$, y c) existe heterocedasticidad, pues $V(Y_j) = \pi_i(1-\pi_i)$, y las π_i 's no tienen porque ser iguales.

Como alternativa al modelo lineal antes mencionado, el modelo logístico propone que cierta función del valor esperado de Y es la que se puede expresar como un modelo lineal. El modelo está dado por

$$\eta = \log\left(\frac{\pi}{1-\pi}\right) = \sum_{j=0}^{p-1} \beta_j x_j,$$

en donde $\log(\cdot)$ es la función logaritmo natural. La transformación η es conocida como *logit*(π), y es claro que η transforma el intervalo $(0,1)$ a todo \mathfrak{R} .

En términos de las observaciones, el modelo logístico será entonces

$$\eta = X \beta,$$

donde $\eta^T = (\eta_1, \eta_2, \dots, \eta_n)$ y $\beta^T = (\beta_0, \beta_1, \dots, \beta_{p-1})$.

Implícitamente, el modelo logístico supone que la probabilidad de "éxito" se puede expresar como

$$\pi = \frac{\exp\left\{\sum_{j=0}^{p-1} \beta_j x_j\right\}}{1 + \exp\left\{\sum_{j=0}^{p-1} \beta_j x_j\right\}} = \frac{1}{1 + \exp\left\{-\sum_{j=0}^{p-1} \beta_j x_j\right\}}$$

Es evidente que π siempre estará en el intervalo (0,1), para todo valor del vector de parámetros β y del vector de variables explicativas $\mathbf{x}^T = (x_0, x_1, \dots, x_{p-1})$. Al considerar un término de error definido como la diferencia entre Y y su valor esperado, se obtiene la expresión

$$Y = \pi + \varepsilon = \pi(\mathbf{x}) + \varepsilon,$$

en donde

$$\pi = \pi(\mathbf{x}) = \frac{1}{1 + \exp\left\{-\sum_{j=0}^{p-1} \beta_j x_j\right\}}$$

El término de error es una variable aleatoria con las siguientes características,

$$E(\varepsilon) = 0, \quad V(\varepsilon) = \pi(\mathbf{x}) [1 - \pi(\mathbf{x})] ;$$

lo cual se puede comprobar fácilmente al darse cuenta que el error solamente puede tomar uno de dos posibles valores, $1 - \pi(\mathbf{x})$ y $-\pi(\mathbf{x})$, con probabilidades respectivas dadas por $\pi(\mathbf{x})$ y $1 - \pi(\mathbf{x})$

Modelos alternativos al logístico son el probit, log-log, log-log complementario y angular. Cada uno de estos modelos se basa en la siguiente transformación de $\pi = E(Y)$,

<u>modelo</u>	<u>transformación</u>
probit	$\Phi^{-1}(\pi)$
log-log	$\log\{-\log(1-\pi)\} = F_M^{-1}(\pi)$
log-log complementario	$-\log\{-\log(\pi)\} = F_m^{-1}(\pi)$
angular	$\text{sen}^{-1}(\pi^{1/2})$

La función Φ es la distribución acumulada de una $N(0,1)$, F_M y F_m son las funciones de distribución acumuladas del máximo y del mínimo valor extremo respectivamente.

El uso de la función logística presenta varias ventajas sobre las otras alternativas, tanto desde el punto de vista teórico como práctico. Comparaciones entre estos modelos se pueden encontrar en McCullagh y Nelder (1989), Cox (1989), Chambers y Cox (1967), entre otros.

2.2 MODELOS LINEALES GENERALIZADOS

Mucho del trabajo sobre el modelo logístico se ha desarrollado considerando que este modelo pertenece a la clase de los modelos lineales generalizados. A continuación se describen estos modelos y se plantearán los supuestos que dan lugar al modelo logístico.

Consideremos un conjunto Y_1, Y_2, \dots, Y_n de variables aleatorias independientes. Son tres los elementos que permiten definir a un modelo lineal generalizado: el componente aleatorio, el componente sistemático y la función liga. A continuación se describe cada uno de estos elementos.

- **Componente aleatorio:** la distribución de cada respuesta Y pertenece a la familia exponencial, es decir, la función de densidad de Y se puede expresar como

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

A la constante ϕ se le denomina parámetro de *dispersión*, y si éste es conocido, θ_i es el llamado parámetro *canónico*. No es complicado demostrar que para cada Y_i se cumple lo siguiente,

$$\mu_i = E(Y_i) = b'(\theta_i) \quad , \quad \sigma_i^2 = V(Y_i) = b''(\theta_i) a(\phi) .$$

- **Componente sistemático:** este componente surge al considerar el llamado *predictor lineal* $\eta^T = (\eta_1, \eta_2, \dots, \eta_n)$, dado por

$$\eta_i = \sum_{j=0}^{p-1} \beta_j x_{ij} \quad , \quad i = 1, 2, \dots, n$$

o bien como

$$\eta = X \beta \quad ,$$

en donde X es la matriz formada por las observaciones de las covariables $x_{0i}, x_{1i}, \dots, x_{p-1i}$ y $\beta^T = (\beta_0, \beta_1, \dots, \beta_{p-1})$ es el vector de parámetros.

- **Función liga:** la relación que se propone entre el valor esperado de cada Y y el predictor lineal η queda expresada en una función g llamada *función liga*. Sea entonces

$$g(\mu_i) = g(E(Y_i)) = \eta_i = \sum_{j=0}^{p-1} \beta_j x_{ij} .$$

El parámetro canónico θ_i depende del valor esperado μ_i , pues $\mu_i = b'(\theta_i)$. Si se elige la función liga como $g(\mu_i) = \theta_i$, se tiene la llamada liga canónica, y su uso tiene la ventaja de que asegura la existencia de una estadística suficiente al estimar β .

La obtención del modelo logístico como caso particular de los modelos lineales generalizados es inmediata al considerar lo siguiente,

$$Y_i = \text{bin} \sim (m_i, \pi_i)$$

$$g(\mu_i) = \log\left(\frac{\mu_i}{m_i - \mu_i}\right) = \log\left(\frac{\pi_i}{1 - \pi_i}\right),$$

para todo $i=1,2,\dots,n$. La función liga resulta ser la canónica, como se hace ver en la parte final de esta sección.

El estimador de máxima verosimilitud para el vector β se puede obtener al aplicar iterativamente el procedimiento de mínimos cuadrados ponderados. La ecuación a resolver es

$$\frac{\partial}{\partial \beta} L(\beta) = 0$$

donde $L(\beta)$ es el logaritmo natural de la función de verosimilitud, es decir,

$$L(\beta) = \log \prod_{i=1}^n f(y_i) = \sum_{i=1}^n \log(f(y_i)) = \sum_{i=1}^n l_i(\beta).$$

La función $l_i(\beta)$ es el logaritmo de la verosimilitud para β habiendo observado y_i . Como $f(y)$ pertenece a la familia exponencial, $L(\beta)$ se puede expresar como

$$L(\beta) = \sum_1^n l_i(\beta) = \sum_1^n \{ a^{-1}(\phi) [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \}.$$

Por otra parte, aplicando la regla de la cadena, tenemos que

$$\frac{\partial l_i(\beta)}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j},$$

pero $b'(\theta_i) = \mu_i$ y $\partial \mu_i / \partial \theta_i = \partial b'(\theta_i) / \partial \theta_i = b''(\theta_i)$, así que

$$\frac{\partial l_i(\beta)}{\partial \beta_j} = a^{-1}(\phi) (y_i - \mu_i) [b''(\theta_i)]^{-1} (\partial \eta_i / \partial \mu_i)^{-1} x_{ij}.$$

Sumando sobre todas las observaciones se tiene que

$$\frac{\partial L(\beta)}{\partial \beta_s} = \sum_1^n \frac{\partial l_i(\beta)}{\partial \beta_s} = \sum_1^n (y_i - \mu_i) a^{-1}(\phi) [b''(\theta_i) (\partial \eta_i / \partial \mu_i)]^{-1} x_{ij},$$

por lo que finalmente

$$\frac{\partial L(\beta)}{\partial \beta} = a^{-1}(\phi) \mathbf{X}^T (\mathbf{VC})^{-1} (\mathbf{y} - \boldsymbol{\mu}) \quad (\text{A})$$

en donde $\mathbf{V} = \text{diag}\{b''(\theta_i)\}$, $\mathbf{C} = \text{diag}\{\partial \eta_i / \partial \mu_i\}$. Si se denota con $h(\beta)$ a $\partial L(\beta) / \partial \beta$ (*), el estimador máximo verosímil de β está dado por la solución de la ecuación $h(\beta) = 0$. Como puede verse en (A), esta última ecuación es no lineal en β , por lo que se ha sugerido la aplicación del método de Newton-Raphson, el cual nos dice que en la iteración k la solución para β está dada por

$$\beta^{(k)} = G(\beta^{(k-1)}) \quad (\text{C})$$

(*) A la función $h(\beta)$ se le denomina función de puntaje (score)

en donde

$$G(\beta) = \beta - J^{-1}(\beta) h(\beta) \quad (B)$$

con $J(\beta)$ la matriz Jacobiana cuya entrada (r,s) es igual a

$$\frac{\partial h_r(\beta)}{\partial \beta_s} = \frac{\partial^2 L(\beta)}{\partial \beta_r \partial \beta_s} \quad (C)$$

La sustitución de (B) en (C) nos lleva a la expresión

$$\beta^{(k)} = \beta^{(k-1)} - J^{-1}(\beta^{(k-1)}) h(\beta^{(k-1)}) \quad (E)$$

y al aplicar iterativamente esta relación se obtiene la aproximación del estimador máximo verosímil de β .

Una modificación al método descrito anteriormente propone utilizar el valor esperado de la matriz J . En realidad, es fácil darse cuenta, a partir de (D), que el negativo de este valor esperado es la matriz de información de Fisher, $i(\beta)$, esto es,

$$i(\beta) = -E\{J(\beta)\}$$

El uso de $E\{J(\beta)\}$ en lugar de $J(\beta)$ permite expresar la relación en (E) de la siguiente forma

$$i(\beta^{(k-1)}) \beta^{(k)} = i(\beta^{(k-1)}) \beta^{(k-1)} + h(\beta^{(k-1)}) \quad (F)$$

El elemento (r,s) de la matriz $i(\beta)$ está dado por (ver Apéndice No. 1)

$$a^{-1}(\phi) [b''(\theta)]^{-1} (\partial \eta_i / \partial \theta_j)^2 x_{is} x_{ir}$$

por lo que $i(\beta)$ se puede escribir de la siguiente forma,

$$i(\beta) = a^{-1}(\phi) X (CVC)^{-1} X'$$

de acuerdo a la definición de C y V que se dió en (A).

La sustitución de $i(\beta)$ y de $h(\beta)$ en (F) nos dá finalmente la siguiente relación de recurrencia

$$X^T (C^{(k-1)} V^{(k-1)} C^{(k-1)})^{-1} X \beta^{(k)} = X^T (C^{(k-1)} V^{(k-1)} C^{(k-1)})^{-1} [X \beta^{(k-1)} + C^{(k-1)} (y - \mu^{(k-1)})]. \quad (G)$$

Esta última expresión es la que se hubiera obtenido al aplicar mínimos cuadrados ponderados al modelo

$$Z^{(k-1)} = X \beta^{(k)} + \varepsilon^{(k-1)}$$

con

$$E(\varepsilon^{(k-1)}) = 0, \quad V(\varepsilon^{(k-1)}) = C^{(k-1)} V^{(k-1)} C^{(k-1)}$$

en donde $Z^{(k-1)}$ es igual a

$$Z^{(k-1)} = X \beta^{(k-1)} + C^{(k-1)} (y - \mu^{(k-1)}).$$

Cuando se usa la liga canónica resulta que $\theta_i = g(\mu_i) = \eta_i$, de donde

$$b''(\theta_i) \frac{\partial \eta_i}{\partial \mu_i} = \frac{\partial \mu_i}{\partial \theta_i} \cdot \frac{\partial \eta_i}{\partial \mu_i} = \frac{\partial \eta_i}{\partial \theta_i} = 1$$

por lo que las ecuaciones de verosimilitud serán, por lo establecido en (A),

$$X^T (y - \mu) = 0$$

Adicionalmente, cuando se usa la liga canónica, se puede establecer la existencia de una estadística suficiente para β ; a saber, $X^T y$.

Para finalizar esta sección veremos algunos resultados específicos para el caso del modelo logístico.

El modelo logístico considera que la variable respuesta Y_i es tal que $Y_i \sim \text{bin}(m_i, \pi_i)$. Esta distribución pertenece a la familia exponencial definida en (H), como puede comprobarse al considerar

$$\phi = 1, \quad a(\phi) = 1, \quad b(\theta_i) = m_i \log(1 + e^{\theta_i}), \quad \theta_i = \log \frac{\pi_i}{1 - \pi_i}, \quad c(y_i, \phi) = \binom{m_i}{y_i}.$$

Además, es fácil ver que $\mu_i = b'(\theta_i) = m_i e^{\theta_i} / (1 + e^{\theta_i}) = m_i \pi_i$, de donde la liga canónica está dada por

$$\theta_i = \log \frac{\mu_i}{m_i - \mu_i} = \log \frac{\pi_i}{1 - \pi_i}.$$

El uso de esta función liga nos lleva a la expresión más conocida del modelo logístico,

$$\log \frac{\pi_i}{1 - \pi_i} = \sum_{j=0}^{p-1} \beta_j x_{ij}.$$

Como ya se demostró anteriormente, las ecuaciones de verosimilitud son,

$$X^T(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0},$$

pero ahora, $\boldsymbol{\mu}^T = (m_1 \pi_1, m_2 \pi_2, \dots, m_n \pi_n)$.

Finalmente, la relación dada en (E) para la estimación del vector $\boldsymbol{\beta}$ sería

$$\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^{(k-1)} - [X^T W(\boldsymbol{\beta}^{(k-1)}) X]^{-1} X^T (\mathbf{y} - \boldsymbol{\mu}^{(k-1)}),$$

donde

$$W(\beta^{(k-1)}) = \text{diag} \{ m_i \pi_i (1 - \pi_i) \} ,$$

y

$$\mu_i^{(k-1)} = m_i \pi_i^{(k-1)} = m_i \frac{\exp\{x_i^T \beta^{(k-1)}\}}{1 + \exp\{x_i^T \beta^{(k-1)}\}} .$$

Sobre la distribución del estimador $\hat{\beta}$ se puede decir que tiene una distribución asintótica normal, así que, cuando la muestra es suficientemente grande,

$$E(\hat{\beta}) \approx \beta , \quad \text{COV}(\hat{\beta}) \approx [X(CVC)^{-1}X]^{\dagger}$$

Una discusión formal sobre las condiciones que son necesarias para lograr esta normalidad asintótica se pueden ver en Fahrmeir y Kaufmann (1985). En términos prácticos, estas condiciones se reducen a pedir que $n \rightarrow \infty$ para datos no agrupados (en este caso $m_i = 1, i=1,2,\dots,n$), y para datos agrupados, con n y p fijos, que cada $m_i \rightarrow \infty$.

CAPITULO 3

ESTADISTICA TIPO DIVERGENCIA

3.1 DATOS AGRUPADOS Y NO AGRUPADOS

Es muy conveniente que desde este momento se establezca la diferencia entre tener datos agrupados o no agrupados, pues según sea el caso, es factible aplicar ciertos métodos estadísticos y además, los resultados asintóticos no son los mismos. De acuerdo a lo expuesto en el capítulo anterior, el modelo logístico considera como variables respuesta a Y_1, Y_2, \dots, Y_n , en donde, además de suponer independencia, se tiene que $Y_i \sim \text{bin}(m_i, \pi_i)$, $i=1, 2, \dots, n$. Diremos que los datos no están agrupados si $m_i = 1$, $i=1, 2, \dots, n$. Esto último es lo que ocurre generalmente cuando todas o algunas de las covariables son continuas y también cuando no se tiene el suficiente control en los posibles valores de las covariables, como es el caso en estudios de tipo observacional. Siendo \mathbf{x} el vector de covariables, cada dato (y_i, \mathbf{x}_i) , $i=1, 2, \dots, n$, con información no agrupada, corresponderá a la observación de un solo elemento de la población condicionado a un valor específico del vector de covariables \mathbf{x} .

Los datos están agrupados cuando $m_i > 1$ por lo menos para alguna i . Esta situación puede darse con mayor frecuencia en estudios experimentales, pues se tiene control sobre las covariables, también se da cuando éstas últimas son variables indicadoras asociadas a variables categóricas. Alternativamente, se puede pensar que uno cuenta con datos agrupados cuando se tienen repeticiones (mismo valor de \mathbf{x}) en un conjunto de datos en donde originalmente no se tenían agrupaciones. Teóricamente, cuando las covariables son continuas, es imposible tener repeticiones, no obstante, la resolución que se utilice en las observaciones puede generar la formación de grupos.

Varios de los resultados distribucionales que se presentarán mas adelante son de tipo asintótico, y es conveniente decir que en este sentido se consideran dos situaciones,

- a) cuando el número de observaciones en cada grupo es "grande", lo cual se expresa diciendo que $m_i \rightarrow \infty$, $i = 1, 2, \dots, n$, con n fijo;
- b) cuando se cuenta con "muchas" observaciones clasificadas en un "gran" número de grupos, es decir, cuando $N = \sum_{i=1}^n m_i \rightarrow \infty$ y $n \rightarrow \infty$.

A los resultados con el primer tipo de limite los llamaremos **m-asintóticos**, y a los del segundo tipo **n-asintóticos**.

El análisis de un modelo logístico se facilita si consideramos la siguiente tabla de contingencia, en donde se presentan los valores observados de la variable respuesta para cada valor del vector de covariables x .

Número de observaciones por grupo y categoría

		Grupos				
		x_1	x_2	...	x_n	
Categoría	C_1	y_1	y_2	...	y_n	y_1
	C_2	$m_1 - y_1$	$m_2 - y_2$...	$m_n - y_n$	$N - y_1$
		m_1	m_2	...	m_n	N

Total de elementos : $N = \sum_{i=1}^n m_i$.

Tabla 3.1

Se supondrá un muestreo binomial-producto, es decir, los conteos $Y_i^T = (Y_i, m_i - Y_i)$, $i=1, 2, \dots, n$, son independientes con $Y_i \sim \text{bin}(m_i, \pi_i)$. Además, la

probabilidad en cada celda depende del valor de las covariables, en particular, para el modelo logístico el parámetro π_i está dado por

$$\pi_i = [1 + \exp(-x_i^T \beta)]^{-1}, \quad \beta \in \mathfrak{R}^p.$$

3.2 ESTADÍSTICA TIPO DIVERGENCIA

La mayoría de las pruebas de bondad de ajuste que se han propuesto en el caso de los modelos logísticos, se basan fundamentalmente en el uso de la estadística X^2 de Pearson (Pearson, 1900; Fisher, 1924) o bien en el logaritmo de la razón de verosimilitudes, G^2 . Con el fin de tener resultados más generales, se mencionará en esta sección una familia de estadísticas, presentada inicialmente por Cressie y Read (1984), la cual se construye a partir del concepto de divergencia. Esta familia incluye, entre otras, a las estadísticas X^2 y G^2 .

Cressie y Read proponen, como medida de discrepancia entre valores observados, O_i , y valores esperados, E_i , a la estadística I^λ , dada por,

$$I^\lambda = \frac{1}{N \lambda (\lambda + 1)} \sum_{i=1}^k O_i [(O_i / E_i)^\lambda - 1],$$

en donde los valores esperados se obtienen suponiendo verdadera la hipótesis nula de que la información fué generada por un modelo multinomial con k posibles categorías (en este trabajo $k = 2$); λ es un valor real que define a cada elemento de la familia, y N es el tamaño total de la muestra. Por facilidad, resulta más conveniente multiplicar a la estadística I^λ por $2N$, así que la familia de estadísticas que se considerará de ahora en adelante estará dada por $\{(2N) I^\lambda : \lambda \in \mathfrak{R}\}$, en donde,

$$(2N) I^\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^k O_i [(O_i / E_i)^\lambda - 1]$$

Algunas propiedades de esta familia se verán mas adelante, por el momento basta mencionar que cuando $O_i = E_i$, $i=1,2,\dots,k$, entonces $(2N) I^\lambda = 0$, $\forall \lambda \in \mathfrak{R}$, y además, para cada $\lambda \in \mathfrak{R}$, $(2N) I^\lambda$ crece si los valores observados y esperados discrepan más entre sí. Por otra parte, cada valor de λ define una forma específica para la estadística, como puede verse a continuación,

λ	$(2N) I^\lambda$
0	$G^2 = 2 \sum_{i=1}^k O_i \log(O_i / E_i)$
1	$X^2 = \sum_{i=1}^k (O_i - E_i)^2 / E_i$
-1	$GM^2 = 2 \sum_{i=1}^k E_i \log(E_i / O_i)$
-2	$NM^2 = \sum_{i=1}^k (O_i - E_i)^2 / O_i$
-1/2	$F^2 = 4 \sum_{i=1}^k (\sqrt{O_i} - \sqrt{E_i})^2$

Cuando $\lambda = 0$ y $\lambda = -1$, se utiliza el valor límite de $(2N) I^\lambda$. En ambos casos se aplica el resultado $\log(t) = \lim_{h \rightarrow 0} (t^h - 1) / h$.

Las estadísticas más conocidas son la X^2 de Pearson y la estadística G^2 basada en la razón de verosimilitudes. GM^2 se conoce como la estadística G^2 modificada (Kullback, 1959), NM^2 es la estadística X^2 modificada (Neyman, 1949) y F^2 es la estadística Freeman-Tukey (Freeman-Tukey, 1950).

En general, un mal ajuste se verá reflejado en un valor grande de $(2N) I^{\lambda}$, pero ¿cómo determinar cuándo un valor de $(2N) I^{\lambda}$ es grande?. Para responder a esta pregunta será necesario conocer la distribución de $(2N) I^{\lambda}$ bajo la hipótesis de nulidad. En este trabajo, la hipótesis nula a considerar es la que nos afirma que un modelo logístico es el adecuado para explicar el fenómeno bajo estudio. La obtención de la distribución exacta de $(2N) I^{\lambda}$ es un trabajo muy complicado, por lo que hay que recurrir a resultados asintóticos. En la siguiente sección se presentan este tipo de resultados.

3.3 RESULTADOS ASINTOTICOS

En esta sección se presenta la distribución asintótica de la estadística $(2N) I^{\lambda}$. Los supuestos generales que se harán son:

- (a1) El tamaño total de la muestra tiende a infinito, es decir, $N \rightarrow \infty$.
- (a2) La dimensión del espacio de covariables, p , permanece fija.
- (a3) El número de categorías, k , para la variable respuesta permanece fijo. En este caso $k=2$, a menos que se especifique otra cosa.

Se considerarán dos casos. En el primero se utilizarán los supuestos clásicos para celdas fijas, es decir, se supone que:

- (b1) El número de grupos (distintos valores del vector de covariables), permanece fijo, esto es, n es fijo, y por lo tanto también el número de celdas.
- (b2) El número de observaciones por grupo tiende a infinito, es decir, $m_i \rightarrow \infty$, $i = 1, 2, \dots, n$, lo cual implica, junto con (b1), que los valores esperados también tienden a infinito.

Los resultados **m-asintóticos** son precisamente los que se obtienen al suponer (a1), (a2), (a3), (b1) y (b2).

En el segundo caso a considerar se supone que:

- (c1) El número de grupos, y por lo tanto el número de celdas, se incrementa, esto es, $n \rightarrow \infty$.

Los resultados obtenidos bajo los supuestos (a1), (a2), (a3) y (c1) son los denominados **n-asintóticos**. En realidad, (c1) es un supuesto de dispersión, en el sentido de que la tabla 3.1 tendrá valores observados cada vez más pequeños.

3.3.1 RESULTADOS m-ASINTOTICOS

El resultado que se presenta a continuación muestra cómo realizar una prueba de bondad de ajuste a partir de la estadística (2N) I^* . El modelo que se supone ha generado la información es el logístico.

TEOREMA 1

Sean Y_1, Y_2, \dots, Y_n , variables aleatorias independientes tales que $Y_i \sim \text{bin}(m_i, \pi_i)$, $i=1, 2, \dots, n$. Consideremos la hipótesis $H_0: \pi_i = F(x_i, \beta)$, en donde $\beta \in \mathfrak{R}^p$ es un vector de parámetros, $x_i \in \mathfrak{R}^p$ es el vector observado de covariables, y $F(x_i, \beta)$ es una función dada por

$$F(x_i, \beta) = G(x_i^T \beta) = \frac{1}{1 + \exp\{-x_i^T \beta\}}$$

Bajo ciertas condiciones de regularidad, es posible asegurar que, cuando H_0 es verdadera, la estadística (2N) I^* converge en distribución, cuando $m_i \rightarrow \infty$, $i=1, 2, \dots, n$, a una variable aleatoria Ji-cuadrada central con $n - p$ grados de libertad.

A continuación se presenta un bosquejo de la demostración del Teorema 1. Primero se supone que $n = 1$ y que la hipótesis nula especifica totalmente el valor

de π_1 , así que la hipótesis H_0 es simple. En este caso la tabla 3.1 se reduce a lo siguiente,

Número de observaciones en el grupo por categoría

		Grupo Único
		x_1
Categoría	C_1	y_1
	C_2	$m_1 - y_1$
		m_1

Total de elementos : $N = m_1$.

Sea U_N la variable aleatoria definida como $U_N = N^{1/2} [(1/N) Y - \pi_0]$, en donde Y y π_0 están dados por

$$Y^T = (Y_1, m_1 - Y_1) \quad , \quad \pi_0^T = (\pi_0, 1 - \pi_0) \quad , \quad \rightarrow$$

y π_0 es el valor del parámetro π_1 , especificado por la hipótesis nula ($H_0: \pi_1 = \pi_0$). La función generatriz de momentos de U_N , $m_N(t)$, está dada por

$$m_N(t) = [\pi_0 \exp\{N^{1/2}(t_1 - t^T \pi_0)\} + (1 - \pi_0) \exp\{N^{1/2}(t_2 - t^T \pi_0)\}]^N$$

Al desarrollar en series la función exponencial, esta última función se puede expresar como

$$m_N(t) = \left\{ 1 - \frac{t^T (D_0 - \pi_0 \pi_0^T) t}{2N} - o_p(N^{-1}) \right\}^N \quad , \quad \text{con } D_0 = \begin{pmatrix} \pi_0 & 0 \\ 0 & 1 - \pi_0 \end{pmatrix}.$$

El término $o(N^{-1})$ converge a cero en probabilidad, así que cuando N tiende a infinito, resulta que

$$m_N(\mathbf{t}) = \exp\left\{\frac{1}{2}\mathbf{t}^T(D_0 - \pi_0\pi_0^T)\mathbf{t}\right\}.$$

A partir de esto último resulta evidente que cuando $N \rightarrow \infty$,

$$\mathbf{U}_N \xrightarrow{d} \mathbf{U} \sim N_2(\mathbf{0}, \pi_0\pi_0^T)$$

Por otra parte, si multiplicamos D_0^{-1} por la matriz de covarianzas de \mathbf{U} , se obtiene una matriz idempotente, y esto nos permite asegurar (ver Searle, 1971, p. 57) que la forma cuadrática $\mathbf{U}^T D_0^{-1} \mathbf{U}$ tiene una distribución $\chi^2_{(1)}$, en donde los grados de libertad están dados por el rango de D_0^{-1} .

Por lo anteriormente expuesto se tiene que, cuando $N \rightarrow \infty$,

$$Q_1 = \mathbf{U}_N^T D_0^{-1} \mathbf{U}_N \xrightarrow{d} \mathbf{U}^T D_0^{-1} \mathbf{U} \sim \chi^2_{(1)}.$$

Ahora bien, la forma cuadrática $Q_1 = \mathbf{U}_N^T D_0^{-1} \mathbf{U}_N$ resulta ser la ya conocida estadística χ^2 de Pearson (Pearson, 1900) pues

$$\chi^2 = \frac{(Y_1 - N\pi_0)^2}{N\pi_0} + \frac{(N - Y_1 - N(1 - \pi_0))^2}{N(1 - \pi_0)} = \mathbf{U}_N^T D_0^{-1} \mathbf{U}_N.$$

Como se mencionó anteriormente, la estadística χ^2 de Pearson es un caso particular de la estadística $(2N)I^1$ cuando $\lambda = 1$, así que

$$(2N)I^1 \sim \chi^2,$$

cuando $N \rightarrow \infty$.

La generalización para toda $\lambda \in \mathfrak{R}$ es inmediata si se demuestra que

$$(2N) I^\lambda = (2N) I^1 + o_p(1) \quad , \quad \lambda \in \mathfrak{R}.$$

Esto último se deduce al expresar $(2N) I^\lambda$ en términos de $V_1 = (Y_1 - N\pi_0)/N\pi_0$ y $V_2 = (N - Y_1 - N(1 - \pi_0)) / N(1 - \pi_0)$, y aplicar una expansión en series de Taylor a $(1 + V_i)^{\lambda-1}$, $i = 1, 2$,

$$\begin{aligned} (2N) I^\lambda &= \frac{2}{\lambda(\lambda+1)} Y_1 \left[\left(\frac{Y_1}{N\pi_0} \right)^\lambda - 1 \right] + \frac{2}{\lambda(\lambda+1)} (N - Y_1) \left[\left(\frac{N - Y_1}{N(1 - \pi_0)} \right)^\lambda - 1 \right] = \\ &= \frac{2N}{\lambda(\lambda+1)} \pi_0 \left[\left(1 + (Y_1 - N\pi_0) / N\pi_0 \right)^{\lambda-1} - 1 \right] + \\ &\quad + \frac{2N}{\lambda(\lambda+1)} (1 - \pi_0) \left[\left(1 + (N - Y_1 - N(1 - \pi_0)) / N(1 - \pi_0) \right)^{\lambda-1} - 1 \right] = \\ &= \frac{2N}{\lambda(\lambda+1)} \pi_0 \left[(1 + V_1)^{\lambda-1} - 1 \right] + \frac{2N}{\lambda(\lambda+1)} (1 - \pi_0) \left[(1 + V_2)^{\lambda-1} - 1 \right] = \\ &= N\pi_0 V_1^2 + N(1 - \pi_0) V_2^2 + o(N^{-1}) \end{aligned}$$

Finalmente, al sustituir V_1 y V_2 , se tiene el resultado deseado,

$$(2N) I^\lambda = (2N) I^1 + o_p(1) \quad , \quad \lambda \in \mathfrak{R}.$$

(Cuando $\lambda = 0$ o $\lambda = -1$, se llega al mismo resultado, solo que hay que calcular el límite de $(2N) I^\lambda$ cuando λ tiende a 0 y cuando λ tiende a -1 respectivamente. El resultado límite que se utiliza es: $\log(t) = \lim_{h \rightarrow 0} (t^h - 1)/h$).

Se recuerda que todo lo hecho anteriormente fué para el caso de tener un solo grupo ($n = 1$) y una hipótesis nula simple. ¿Qué ocurre si hay más de un grupo?. Así como la forma cuadrática $Q_1 = U_1^T D_0^{-1} U_1$, construida a partir de Y_1 , converge en distribución a una $\chi^2_{(1)}$, cuando $N = m_1 \rightarrow \infty$, lo mismo ocurre con las

formas cuadráticas Q_2, Q_3, \dots, Q_n correspondientes a Y_2, Y_3, \dots, Y_n . Por otra parte, en el enunciado del Teorema 1 se dijo que Y_1, Y_2, \dots, Y_n son variables aleatorias independientes, así que cuando $m_i \rightarrow \infty, i = 1, 2, \dots, n$, es posible concluir que $(2N) I^\lambda = \sum_n Q_i$ converge en distribución a una $\chi^2_{(n)}$.

Para terminar con la demostración del Teorema 1, veamos ahora cuál es la distribución asintótica de $(2N) I^\lambda$ cuando la hipótesis nula es compuesta y específica que cada probabilidad π_i depende del vector de covariables x_i y de un vector de parámetros $\beta \in \mathfrak{R}^p$. La hipótesis nula es compuesta porque el vector β es desconocido. En este caso, el cálculo de la estadística X^2 requiere de la estimación de β , pues los valores esperados se obtienen utilizando las probabilidades $\hat{\pi}_i = G(x_i^T \hat{\beta}), i = 1, 2, \dots, n$. La expresión de la estadística X^2 es la siguiente,

$$X^2 = \sum_1^n \left[\frac{(Y_i - m_i \hat{\pi}_i)^2}{m_i \hat{\pi}_i} + \frac{(m_i - Y_i - m_i(1 - \hat{\pi}_i))^2}{m_i(1 - \hat{\pi}_i)} \right].$$

La comparación que se hace con esta X^2 entre valores observados y esperados, corresponde a la siguiente tabla,

Número de observaciones por grupo y categoría

		Grupos				
		x_1	x_2	\dots	x_n	
Categoría	C_1	y_1	y_2	\dots	y_n	$y \cdot$
	C_2	$m_1 - y_1$	$m_2 - y_2$	\dots	$m_n - y_n$	$N - y \cdot$
		m_1	m_2	\dots	m_n	N

Total de elementos : $N = \sum_1^n m_i$.

Sea ahora $U_i = \sqrt{m_i} ((1/m_i)Y_i - \hat{\pi}_i)$, en donde $\hat{\pi}_i^T = (\hat{\pi}_i, 1 - \hat{\pi}_i)$ es un estimador consistente de π_i , $i = 1, 2, \dots, n$. Para obtener la distribución asintótica de X^2 , es conveniente expresar a esta estadística en términos de una forma cuadrática,

$$\begin{aligned} X^2 &= \sum_1^n \left[\frac{(Y_i - m_i \hat{\pi}_i)^2}{m_i \hat{\pi}_i} + \frac{(m_i - Y_i - m_i(1 - \hat{\pi}_i))^2}{m_i(1 - \hat{\pi}_i)} \right] = \\ &= \sum_1^n \left\{ \left[\sqrt{m_i} \left(\frac{1}{m_i} Y_i - \hat{\pi}_i \right) \right] \frac{1}{\hat{\pi}_i} \sqrt{m_i} \left(\frac{1}{m_i} Y_i - \hat{\pi}_i \right) \right\} + \\ &\quad + \left\{ \left[\sqrt{m_i} \left(\frac{1}{m_i} (m_i - Y_i) - (1 - \hat{\pi}_i) \right) \right] \frac{1}{1 - \hat{\pi}_i} \sqrt{m_i} \left(\frac{1}{m_i} (m_i - Y_i) - (1 - \hat{\pi}_i) \right) \right\} = \\ &= \sum_1^n U_i^T D_i^{-1} U_i = U_N^T D_N^{-1} U_N \end{aligned}$$

en donde $D_i = \text{diag}\{\hat{\pi}_i, 1 - \hat{\pi}_i\}$, $U_N^T = (U_1^T U_2^T \dots U_n^T)$, y $D_N = \text{diag}\{D_1, D_2, \dots, D_n\}$.

Sea π_i^* el verdadero y desconocido valor de π_i bajo H_0 . Como $\hat{\pi}_i$ es un estimador consistente de π_i^* , resulta que $\hat{\pi}_i = \pi_i^* + o_p(1)$, $i = 1, 2, \dots, n$, y entonces es posible decir que

$$X^2 = U_N^T D_N^{-1} U_N = U_N^T D^{-1} U_N + o_p(1),$$

con $D = \text{diag}\{D_1, D_2, \dots, D_n\}$, $D_i = \text{diag}\{\pi_i^*, 1 - \pi_i^*\}$, $i = 1, 2, \dots, n$.

Las condiciones de regularidad a las que hace referencia el Teorema 1, permiten también establecer que, cuando $m_i \rightarrow \infty$, $i = 1, 2, \dots, n$, U_N^T converge en distribución a una variable aleatoria normal, pues cada U_i^T también converge a

una distribución normal, y además, la dependencia que existe entre las U_i^* 's debido a que cada $\hat{\pi}_i$ depende de $\hat{\beta}$ ($\hat{\pi}_i = G(\mathbf{x}_i, \hat{\beta})$), va desapareciendo conforme cada m_i tiende a infinito. De esto último se desprende que

$$(2N) I^1 = X^2 = \mathbf{U}_N^{*T} D_N^{-1} \mathbf{U}_N^* + o_p(1) \xrightarrow{d} \chi_{(n-p)}^2,$$

en donde los grados de libertad están dados por el rango de la matriz $D \cdot \text{Var}(\mathbf{U}^*)$, con \mathbf{U}^* tal que $\mathbf{U}_N^* \xrightarrow{d} \mathbf{U}^*$.

Finalmente, y como ya se hizo ver anteriormente,

$$(2N) I^\lambda = (2N) I^1 + o_p(1), \quad \lambda \in \mathfrak{R},$$

así que $(2N) I^\lambda$ converge en distribución a una $\chi_{(n-p)}^2$ cuando $m_i \rightarrow \infty$, $i = 1, 2, \dots, n$.

Las condiciones de regularidad a las que se hace referencia en el Teorema 1 están enunciadas en el Apéndice No. 2. Básicamente, estas condiciones piden que la función F , que define la relación entre las probabilidades de éxito, π_i , y un vector de parámetros, β , sea una función suave, que sea positiva, y que, en realidad, el modelo propuesto tenga p parámetros y no menos.

3.3.2 RESULTADOS n -ASINTÓTICOS.

Recordemos que los resultados n -asintóticos son los que se obtienen bajo los supuestos (b1), (b2) y (c1), especificados al inicio de la sección 3.3. Básicamente, estos supuestos se cumplen cuando al incrementar el tamaño total de la muestra, N , también se incrementa el número de grupos, n , pero de tal forma que el número de observaciones en cada grupo, m_i , es relativamente pequeño comparado con N . En esta situación, los valores esperados bajo H_0 no

necesariamente aumentarán conforme se tengan más observaciones. El resultado que se tiene en este caso es el siguiente,

TEOREMA 2

Sean Y_1, Y_2, \dots, Y_n , variables aleatorias independientes tales que $Y_i \sim \text{bin}(m_i, \pi_i)$, $i=1, 2, \dots, n$. Consideremos la hipótesis $H_0 : \pi_i = F(\mathbf{x}_i, \beta)$, en donde $\beta \in \mathfrak{R}^p$ es un vector de parámetros, $\mathbf{x}_i \in \mathfrak{R}^p$ es el vector observado de covariables, y $F(\mathbf{x}_i, \beta)$ es una función dada por

$$F(\mathbf{x}_i, \beta) = G(\mathbf{x}_i^T \beta) = \frac{1}{1 + \exp\{-\mathbf{x}_i^T \beta\}}$$

Bajo ciertas condiciones de regularidad, es posible asegurar que, cuando H_0 es verdadera, la distribución asintótica de $(2N) I^\lambda$, cuando $N \rightarrow \infty$ y $n \rightarrow \infty$, es tal que

$$T_\lambda = \frac{(2N) I^\lambda - \mu_\lambda}{\sigma_\lambda} \sim N(0, 1) \quad , \quad \lambda > -1$$

en donde,

$$\mu_\lambda = E\left\{(2N) I^\lambda(\hat{\beta})\right\} \quad , \quad \sigma_\lambda^2 = V\left\{(2N) I^\lambda(\hat{\beta})\right\} - Q(\hat{\beta}) \quad ,$$

$$Q(\beta) = \mathbf{c}_\lambda^T(\beta) i^{-1}(\beta) \mathbf{c}_\lambda(\beta) \quad ,$$

$$\mathbf{c}_\lambda(\beta) = \text{COV}\left\{(2N) I^\lambda(\beta), \frac{\partial L(\beta)}{\partial \beta}\right\} \quad ,$$

siendo $i(\beta)$ la matriz de información de Fisher y $L(\beta)$ el logaritmo de la función de verosimilitud. La demostración de este teorema en todo detalle puede verse en Rojek (1989). En el Apéndice 3 de este trabajo se presenta un esbozo de la demostración.

Las expresiones para la matriz de información $i(\beta)$ y para el vector de puntajes $\partial L(\beta)/\partial \beta$, fueron establecidas en el capítulo anterior.

La mayoría de los trabajos sobre la distribución n -asintótica de $(2N) I^\lambda$ se han restringido al caso de $\lambda = 1$ (estadística X^2 de Pearson) y $\lambda = 0$ (estadística G^2 de razón de verosimilitudes). Ver los trabajos de Koehler (1986), McCullagh (1985a, 1986), Dale (1986). No fué sino hasta 1992 que Osius y Rojek (Osius y Rojek, 1992) extienden el resultado para toda $\lambda > -1$ (si $\lambda \leq -1$ no se podrían manejar valores observados iguales a cero, como puede verse a partir de la definición de $(2N) I^\lambda$ dada en la sección 3.2). Además, la función G que aparece en el teorema nos define una función lisa G que pudiera ser cualquiera de las mencionadas al final de la sección 2.1. En este trabajo nos interesa en particular el modelo logístico, para el cual $G^{-1}(\pi) = \text{logit}(\pi)$.

Ahora se darán las expresiones adecuadas que nos permitirán en la práctica evaluar μ_λ y σ_λ^2 . Para esto es conveniente reexpresar a la estadística $(2N) I^\lambda$ de la siguiente forma,

$$\begin{aligned} (2N) I^\lambda &= \frac{2}{\lambda(\lambda+1)} \sum_{k=1}^2 \sum_{i=1}^n \left\{ O_{ki} \left[(O_{ki} / E_{ki})^\lambda - 1 \right] \right\} = \\ &= \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^n \left\{ O_{1i} \left[(O_{1i} / E_{1i})^\lambda - 1 \right] + O_{2i} \left[(O_{2i} / E_{2i})^\lambda - 1 \right] \right\} = \\ &= \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^n \left\{ Y_i \left[\left(\frac{Y_i}{m_i \pi_i} \right)^\lambda - 1 \right] + (m_i - Y_i) \left[\left(\frac{m_i - Y_i}{m_i (1 - \pi_i)} \right)^\lambda - 1 \right] \right\} = \\ &= \sum_{i=1}^n A_\lambda(Y_i, m_i, \pi_i) \end{aligned}$$

en donde E_{ki} y O_{ki} son los valores esperados y observados respectivamente en la categoría k del grupo i , y

$$A_{\lambda}(Y, m\pi) = \frac{2}{\lambda(\lambda+1)} Y \left[\left(\frac{Y}{m\pi} \right)^{\lambda} - 1 \right] + \frac{2}{\lambda(\lambda+1)} (m-Y) \left[\left(\frac{m-Y}{m(1-\pi)} \right)^{\lambda} - 1 \right]$$

A continuación se definen algunas cantidades que son de utilidad en desarrollos posteriores,

$$\begin{aligned} e_{\lambda}(m, \pi) &= E \{ A_{\lambda}(Y, m\pi) \}, \\ h_{\lambda}^2(m, \pi) &= V \{ A_{\lambda}(Y, m\pi) \}, \\ c_{\lambda}(m, \pi) &= \text{COV} \{ Y, A_{\lambda}(Y, m\pi) \}. \end{aligned}$$

El valor esperado y la varianza de $(2N) I^{\lambda}$, necesarios para calcular los valores μ_{λ} y σ_{λ}^2 del Teorema 2, están dados por

$$\begin{aligned} \mu_{\lambda}(\beta) &= E\{(2N) I^{\lambda}\} = \sum_{i=1}^n e_{\lambda}(m_i, \pi_i), \\ v_{\lambda}^2(\beta) &= V\{(2N) I^{\lambda}\} = \sum_{i=1}^n h_{\lambda}^2(m_i, \pi_i), \end{aligned}$$

y el vector $c_{\lambda}(\beta)$, que aparece en la forma cuadrática de σ_{λ}^2 es igual a

$$c_{\lambda}(\beta) = X^T \text{diag} \{ G'(x_i^T \beta) [\pi_i (1-\pi_i)] \} c_{\lambda}(m, \pi),$$

donde $c_{\lambda}(m, \pi)$ es el vector cuyo componente i es $c_{\lambda}(m_i, \pi_i)$.

Es un ejercicio sencillo, pero bastante largo, el ver que para la estadística X^2 de Pearson ($\lambda=1$), se tiene

$$\begin{aligned} e_1(m, \pi) &= E \{ A_1(Y, m\pi) \} = 1, \\ h_1^2(m, \pi) &= V \{ A_1(Y, m\pi) \} = 2 + [\pi^{-1}(1-\pi^{-1}) - 6] / m, \end{aligned}$$

$$c_1(m, \pi) = \text{COV}\{Y, A_1(Y, m, \pi)\} = 1 - 2\pi.$$

No obstante, el cálculo de e_λ , v_λ^2 y c_λ para cualquier $\lambda > -1$ requerirá, en general, de la suma sobre todos los posibles valores de la variable de interés en cada grupo. El esfuerzo computacional que se requiere para calcular estas cantidades se incrementa rápidamente conforme aumenta el número de posibles valores para la variable respuesta (muestreo multinomial-producto).

A continuación se presenta otra alternativa para el cálculo de σ_λ^2 , la cual resulta interesante tanto desde el punto de vista práctico como teórico. Se recuerda que $\sigma_\lambda^2 = v_\lambda^2(\beta) - Q(\beta)$, en donde

$$v_\lambda^2(\beta) = V\{(2N) I^\lambda(\beta)\} = \sum_{i=1}^n h_\lambda^2(m_i, \pi_i),$$

$$Q(\beta) = c_\lambda^T(\beta) i^{-1}(\beta) c_\lambda(\beta),$$

$$c_\lambda(\beta) = X^T \text{diag}\{G'(x_i^T \beta) [\pi_i (1 - \pi_i)]\} c_\lambda(m, \pi).$$

Consideremos la regresión lineal múltiple con matriz de diseño X y respuesta dada por

$$u_\lambda(\beta) = \text{diag}^{-1}\{m_i G'(x_i^T \beta)\} c_\lambda(m, \pi).$$

Al aplicar mínimos cuadrados ponderados con matriz de pesos

$$W(\beta) = \text{diag}\{m_i G'(x_i^T \beta)^2 \pi_i^{-1} (1 - \pi_i^{-1})\},$$

se obtiene la siguiente suma de cuadrados de residuales (SCR),

$$\begin{aligned} \text{SCR}(\beta) &= u_\lambda^T(\beta) [W(\beta) - W(\beta) X i^{-1}(\beta) X^T W(\beta)] u_\lambda(\beta) = \\ &= u_\lambda^T(\beta) W(\beta) u_\lambda(\beta) - u_\lambda^T(\beta) W(\beta) X i^{-1}(\beta) X^T W(\beta) u_\lambda(\beta) \end{aligned}$$

($W(\beta)$ es la matriz $(CVC)^{-1}$ definida en el capítulo 2).

A partir de las expresiones de $u_\lambda(\beta)$ y de $W(\beta)$, resulta que

$$u_\lambda^T(\beta) W(\beta) u_\lambda(\beta) = \sum_n [c_\lambda^2 (m_i, \pi_i) m_i^{-1} \pi_i^{-1} (1 - \pi_i^{-1})] ,$$

$$X^T W(\beta) u_\lambda(\beta) = X^T \text{diag} \{ G'(x_i^T \beta) \pi_i^{-1} (1 - \pi_i^{-1}) \} c_\lambda(m, \pi) = c_\lambda(\beta) .$$

Entonces,

$$\begin{aligned} SCR(\beta) &= u_\lambda^T(\beta) W(\beta) u_\lambda(\beta) - c_\lambda^T(\beta) i^{-1}(\beta) c_\lambda(\beta) = \\ &= \sum_n [c_\lambda^2 (m_i, \pi_i) m_i^{-1} \pi_i^{-1} (1 - \pi_i^{-1})] - Q(\beta) , \end{aligned}$$

de donde

$$Q(\beta) = \sum_n [c_\lambda^2 (m_i, \pi_i) m_i^{-1} \pi_i^{-1} (1 - \pi_i^{-1})] - SCR(\beta) .$$

Finalmente, la expresión para σ_λ^2 sería la que se muestra a continuación

$$\begin{aligned} \sigma_\lambda^2 &= v_\lambda^2(\beta) - Q(\beta) = \sum_n h_\lambda^2 (m_i, \pi_i) - \sum_n [c_\lambda^2 (m_i, \pi_i) m_i^{-1} \pi_i^{-1} (1 - \pi_i^{-1})] + SCR(\beta) = \\ &= \sum_n [h_\lambda^2 (m_i, \pi_i) - c_\lambda^2 (m_i, \pi_i) m_i^{-1} \pi_i^{-1} (1 - \pi_i^{-1})] + SCR(\beta) \end{aligned}$$

Sea V_λ la suma

$$V_\lambda = \sum_n [h_\lambda^2 (m_i, \pi_i) - c_\lambda^2 (m_i, \pi_i) m_i^{-1} \pi_i^{-1} (1 - \pi_i^{-1})] .$$

Cada sumando de V_λ es no negativo ya que

$$h_\lambda^2 (m, \pi) - c_\lambda^2 (m, \pi) m^{-1} \pi^{-1} (1 - \pi^{-1}) = V \{ A_\lambda (Y, m\pi) \} - \text{COV}^2 \{ Y, A_\lambda (Y, m\pi) \} V^{-1}(Y) ,$$

y siempre se cumple que

$$0 \leq \frac{\text{COV}^2\{Y, A_\lambda(Y, m\pi)\}}{V(Y) V\{A_\lambda(Y, m\pi)\}} \leq 1 .$$

Cuando $\lambda = 1$ la expresión de V_λ es muy sencilla,

$$V_1 = 2 \left[n - \sum_{i=1}^n m_i^{-1} \right] .$$

3.3.2.1 DISPERSION EXTREMA.

En la práctica, es común encontrar situaciones en donde solamente se tiene una observación dentro de cada uno de los grupos definidos por el vector de covariables. En otras palabras, la variable respuesta, Y_i , es tal que $Y_i \sim \text{bin}(1, \pi_i)$, $i = 1, 2, \dots, n$. Es interesante ver qué ocurre con la distribución n -asintótica de la estadística $(2N) I^\lambda$ en este caso.

En la sección anterior se estableció que la varianza asintótica de $(2N) I^\lambda$ se puede expresar como

$$\sigma_\lambda^2 = V_\lambda + \text{SCR}(\beta) .$$

Como ya se hizo ver anteriormente, V_λ es no negativa para todo valor de λ . Cuando $m_i = 1$, $i = 1, 2, \dots, n$, resulta que V_λ es igual a cero (ver Anexo No.4), por lo que $\sigma_\lambda^2 = \text{SCR}(\beta)$. La suma de cuadrados de residuales, $\text{SCR}(\beta)$, se obtiene al considerar una regresión lineal con matriz de diseño X y respuesta $u_\lambda(\beta)$ dada por $u_\lambda(\beta) = \text{diag}^{-1}\{m_i, G'(x_i^T, \beta)\} c_\lambda(m, \pi)$, así que cuando $u_\lambda(\beta)$ pertenece al subespacio generado por las columnas de la matriz X , es evidente que $\text{SCR}(\beta) = 0$. Esto último es lo que sucede con el modelo logístico cuando $\lambda = 0$, veamos,

$$\begin{aligned}
G'(\mathbf{x}^T \boldsymbol{\beta}) &= G'(\eta) = \frac{d}{d\eta} [1 + \exp(-\eta)]^{-1} = \\
&= [1 + \exp(-\eta)]^{-1} \{1 - [1 + \exp(-\eta)]^{-1}\} = \\
&= G(\eta) [1 - G(\eta)] = \pi(1 - \pi) ,
\end{aligned}$$

y por otra parte (ver Apéndice No.3),

$$c_{\lambda}(1, \pi) = \frac{2\pi(1-\pi)}{\lambda(\lambda+1)} \left[\left(\frac{1}{\pi}\right)^{\lambda} - \left(\frac{1}{1-\pi}\right)^{\lambda} \right] ,$$

entonces, $u_{\lambda}(\boldsymbol{\beta})$ es un vector cuyo elemento i -ésimo está dado por

$$u_{\lambda}(\boldsymbol{\beta})_i = \frac{2}{\lambda(\lambda+1)} \left[\left(\frac{1}{\pi}\right)^{\lambda} - \left(\frac{1}{1-\pi}\right)^{\lambda} \right] ,$$

Para $\lambda = 0$, $u_0(\boldsymbol{\beta})_i$ de la siguiente forma,

$$\begin{aligned}
u_0(\beta)_i &= \lim_{\lambda \rightarrow 0} u_\lambda(\beta)_i = 2 \lim_{\lambda \rightarrow 0} \left[\frac{1}{\lambda} \left(\frac{1}{\pi_i} \right)^\lambda - \frac{1}{\lambda} - \frac{1}{\lambda} \left(\frac{1}{1-\pi_i} \right)^\lambda + \frac{1}{\lambda} \right] = \\
&= 2 \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left[\left(\frac{1}{\pi_i} \right)^\lambda - 1 \right] - 2 \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left[\left(\frac{1}{1-\pi_i} \right)^\lambda - 1 \right] = \\
&= 2 \left[\log \left(\frac{1}{\pi_i} \right) - \log \left(\frac{1}{1-\pi_i} \right) \right] = 2 \log \left(\frac{1-\pi_i}{\pi_i} \right) = -2 \operatorname{logit}(\pi_i) .
\end{aligned}$$

Esto último nos dice que $u_0(\beta)$ pertenece al subespacio generado por las columnas de X , y entonces $\sigma_\beta^2 = \operatorname{SCR}(\beta) = 0$, lo cual corrobora la afirmación de McCullagh (1985), en el sentido de que la estadística $G^2 = (2N) I^0$ no es adecuada para realizar una prueba de bondad de ajuste cuando en cada grupo hay muy pocas observaciones.

La prueba de bondad de ajuste basada en el resultado del Teorema 2, se puede ver como una prueba realizada con una estadística de puntajes, en donde se considera un modelo de la forma

$$\pi = H_\lambda(\mathbf{x}^T \beta) = G(\mathbf{x}^T \beta + \psi h_\lambda(\mathbf{x}^T \beta)) ,$$

con

$$h_\lambda(y) = [A_\lambda(1, G(y)) - A_\lambda(0, G(y))] G(y) [1 - G(y)] / G'(y) , \quad y \in \mathfrak{R} .$$

Al probar la hipótesis nula $H_0 : \psi = 0$, y aplicar el método de puntajes, se llega a la misma estadística T_λ del Teorema 2. Esto sugiere que cuando $m_i = 1$, $i = 1, 2, \dots, n$, la prueba basada en T_λ se puede considerar, mas que como una prueba de

bondad de ajuste global, como una prueba sobre la función ligo propuesta, que en este caso es la transformación *logit*.

CAPÍTULO 4

APLICACION DE RESULTADOS

En este capítulo se dan dos ejemplos en los cuales se aplican los resultados de los Teoremas 1 y 2. El primer ejemplo es de datos agrupados, y el segundo es de datos con extrema dispersión. En ambos casos se calcula el valor de la estadística de prueba para diferentes valores de λ . Los comentarios se harán al final de cada uno de los ejemplos.

4.1 EJEMPLO 1

Los datos de este ejemplo provienen de un estudio realizado para conocer la relación entre la edad de las jóvenes en Varsovia y la probabilidad de que hayan menstruado. Con la variable EDAD se formaron 21 grupos. La marca de clase es el valor que se utiliza en el estudio. La variable respuesta, Y , corresponde al número de mujeres que dijeron haber menstruado.

EDAD	Y	m (total)
10.83	2	120
11.08	2	90
11.33	5	88
11.58	10	105
11.83	17	111
12.08	16	100
12.33	29	93
12.58	39	100
12.83	51	108
13.08	47	99
13.33	67	106
13.58	81	105
13.83	88	117
14.08	79	98
14.33	90	97
14.58	113	120
14.83	95	102
15.08	117	122
15.33	107	111
15.58	92	94
15.83	112	114

La referencia de esta informaciones Guerrero y Johnson (1982).

El ajuste del modelo se muestra a continuación,

LOGIT // Dependent Variable is Y				
Date: 01/10/95 Time: 15:05				
Sample: 1 2200				
Included observations: 2200				
Convergence achieved after 4 iterations				
Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	-20.37206	0.815279	-24.98784	0.0000
EDAD	1.567802	0.062215	25.19959	0.0000
Log likelihood	-813.9078			
Obs with Dep=1	1259			
Obs with Dep=0	941			
Variable	Mean All	Mean D=1	Mean D=0	
C	1.000000	1.000000	1.000000	
EDAD	13.37080	14.30299	12.12357	

Los valores estimados de la probabilidad de éxito, $\hat{\pi}_i$, y de la variable respuesta, son

EDAD	Y	m (total)	$\hat{\pi}_i$	\hat{Y}
10.83	2	120	0.0325	3.9027
11.08	2	90	0.0474	4.2650
11.33	5	88	0.0686	6.0342
11.58	10	105	0.0982	10.3154
11.83	17	111	0.1388	15.4112
12.08	16	100	0.1926	19.2630
12.33	29	93	0.2609	24.2679
12.58	39	100	0.3432	34.3190
12.83	51	108	0.4361	47.0947
13.08	47	99	0.5336	52.8310
13.33	67	106	0.6287	66.6446
13.58	81	105	0.7148	75.0514
13.83	88	117	0.7876	92.1516
14.08	79	98	0.8459	82.8956
14.33	90	97	0.8904	86.3661
14.58	113	120	0.9232	110.7827
14.83	95	102	0.9468	96.5706
15.08	117	122	0.9634	117.5347
15.33	107	111	0.9750	108.2217

15.58	92	94	0.9829	92.3971
15.83	112	114	0.9884	112.6791

La estadística (2N) I^λ se evaluó a partir de su expresión,

$$(2N) I^\lambda = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^n \left\{ Y_i \left[\left(\frac{Y_i}{m_i \hat{\pi}_i} \right)^\lambda - 1 \right] + (m_i - Y_i) \left[\left(\frac{m_i - Y_i}{m_i(1 - \hat{\pi}_i)} \right)^\lambda - 1 \right] \right\}.$$

λ	(2N) I^λ
-2	17.2734
-1.8	16.9803
-1.6	16.7105
-1.4	16.4622
-1.2	16.2339
-1	16.0246
-0.8	15.8319
-0.6	15.6557
-0.4	15.4947
-0.2	15.3478
0	15.2143
0.2	15.0933
0.4	14.9842
0.6	14.8863
0.8	14.7990
1	14.7219
1.2	14.6545
1.4	14.5964
1.6	14.5471
1.8	14.5064
2	14.4740

Los valores de la estadística (2N) I^λ se comparan con el valor en tablas de una $\chi^2_{(21-2)}$ con $n - p = 19$ grados de libertad. Para todo valor de λ no se rechaza la hipótesis de que los datos fueron generados por un modelo logístico. Varias técnicas de diagnóstico fueron aplicadas al mismo conjunto de datos en el trabajo de Rebeca Aguirre (1991), y se corrobora el buen ajuste del modelo.

4.2 EJEMPLO 2

Para este ejemplo se utilizó una muestra de 79 individuos, y lo que se deseaba estudiar era si existía alguna relación entre ciertas propiedades de la orina y la formación de cristales. La información se obtuvo de Andrews et. al. (1985). A continuación se presentan solamente los datos de dos variables: la variable de respuesta, Y, cuyo valor es de 0 si no se encontraron cristales en la orina y de 1 si había cristales; la otra variable es CAL, que representa la concentración de calcio en milmoles por litro.

Y	CAL	Y	CAL
0	2.45	0	5.38
0	4.49	0	3.53
0	2.36	0	4.54
0	2.15	0	3.98
0	1.16	0	1.02
0	3.34	0	3.46
0	1.4	0	1.19
0	8.48	0	5.64
0	1.16	0	2.66
0	2.21	0	1.22
0	1.93	0	2.64
0	1.27	0	2.31
0	1.03	0	4.49
0	1.47	1	6.96
0	1.53	1	13.00
0	5.09	1	5.54
0	1.05	1	6.19
0	2.03	1	7.31
0	7.68	1	14.34
0	1.45	1	4.74
0	5.16	1	2.50
0	0.81	1	1.27
0	1.32	1	4.18
0	1.55	1	3.10
0	1.52	1	3.01
0	0.77	1	6.81
0	2.17	1	8.28
0	0.17	1	2.33
0	0.83	1	7.18
0	3.04	1	5.67
0	1.06	1	12.68
0	3.93	1	8.94

Y	CAL	Y	CAL
1	3.16	1	6.63
1	3.30	1	8.53
1	6.99	1	9.04
1	0.65	1	0.58
1	4.18	1	7.82
1	4.45	1	12.20
1	0.27	1	9.39
1	7.64		

La variable explicativa, CAL, es continua, de tal suerte que para cada uno de sus valores se forma un grupo en donde se tiene una sola observación de la variable respuesta Y. Esto es una situación que en la práctica se presenta con bastante frecuencia. El ajuste del modelo se presenta a continuación,

LOGIT // Dependent Variable is Y				
Date: 01/10/95 Time: 18:29				
Sample: 1 79				
Included observations: 79				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	-2.178392	0.517142	-4.212371	0.0001
CAL	0.468642	0.114495	4.099128	0.0001
Log likelihood	-40.75136			
Obs with Dep=1	34			
Obs with Dep=0	45			
Variable	Mean All	Mean D=1	Mean D=0	
C	1.000000	1.000000	1.000000	
CAL	4.138987	6.142941	2.624889	

El resultado del Teorema 2 es el que se aplica en este caso. El cálculo de la estadística $(2N) I^{\hat{\beta}}$ es semejante al del ejemplo anterior, la única diferencia es que ahora $m_i = 1$, $i = 1, 2, \dots, 79$. Los valores estimados de $\hat{\pi}_i$, necesarios para el cálculo de $(2N) I^{\hat{\beta}}$, μ_i , y σ_i , son

Y	CAL	$\hat{\pi}_1$	Y	CAL	$\hat{\pi}_1$
0	2.45	0.2630	0	5.38	0.5849
0	4.49	0.4815	0	3.53	0.3719
0	2.36	0.2549	0	4.54	0.4873
0	2.15	0.2367	0	3.98	0.4223
0	1.16	0.1632	0	1.02	0.1544
0	3.34	0.3513	0	3.46	0.3643
0	1.4	0.1791	0	1.19	0.1651
0	8.48	0.8576	0	5.64	0.6141
0	1.15	0.1632	0	2.66	0.2829
0	2.21	0.2418	0	1.22	0.1671
0	1.93	0.2186	0	2.64	0.2807
0	1.27	0.1703	0	2.31	0.2505
0	1.03	0.1550	0	4.49	0.4815
0	1.47	0.1840	1	6.96	0.7471
0	1.53	0.1883	1	13.00	0.9804
0	5.09	0.5516	1	5.54	0.6030
0	1.05	0.1563	1	6.19	0.6732
0	2.03	0.2267	1	7.31	0.7768
0	7.68	0.8055	1	14.34	0.9895
0	1.45	0.1826	1	4.74	0.5107
0	5.16	0.5597	1	2.50	0.2676
0	0.81	0.1420	1	1.27	0.1703
0	1.32	0.1737	1	4.18	0.4454
0	1.55	0.1897	1	3.10	0.3282
0	1.52	0.1875	1	3.01	0.3170
0	0.77	0.1397	1	6.81	0.7336
0	2.17	0.2384	1	8.28	0.8458
0	0.17	0.1092	1	2.33	0.2523
0	0.83	0.1431	1	7.18	0.7661
0	3.04	0.3200	1	5.67	0.6175
0	1.06	0.1569	1	12.68	0.9773
0	3.93	0.4166	1	8.94	0.8820
1	3.16	0.3324	1	6.63	0.7168
1	3.30	0.3471	1	8.53	0.8605
1	6.99	0.7498	1	9.04	0.8868
1	0.65	0.1331	1	0.58	0.1294
1	4.18	0.4454	1	7.82	0.8155
1	4.45	0.4768	1	12.20	0.9718
1	0.27	0.1139	1	9.39	0.9022
1	7.64	0.8025			

La expresión de μ_λ es

$$\mu_\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{79} \left\{ \left[\left(\frac{1}{\pi_i} \right)^\lambda - 1 \right] \pi_i + \left[\left(\frac{1}{1-\pi_i} \right)^\lambda - 1 \right] (1-\pi_i) \right\}$$

Como solo hay una observación por grupo, σ_{λ}^2 es igual la suma de cuadrados de residuales de la regresión ponderada que tiene, como variable respuesta, al vector u_{λ} , con elemento i -ésimo igual a

$$\frac{2}{\lambda(\lambda+1)} \left[\left(\frac{1}{\hat{\pi}_i} \right)^{\lambda} - \left(\frac{1}{1-\hat{\pi}_i} \right)^{\lambda} \right],$$

la matriz de diseño es igual a la del modelo logístico, y la matriz de pesos es $W = \text{diag} \{ \hat{\pi}_i, (1 - \hat{\pi}_i) \}$.

A continuación se presenta una tabla con todos los resultados anteriores para distintos valores de λ .

λ .	$(2N) I^{\lambda}$	μ_{λ}	$\sigma_{\lambda}^2 = SCR$	T_{λ}
-0.8	275.9363	289.2091	432.9484	-0.6379
-0.6	149.0699	155.9449	151.3177	-0.5589
-0.4	108.0978	112.8157	95.7735	-0.4821
-0.2	88.8588	92.4692	78.3250	-0.4079
0.2	73.1072	75.5312	79.3154	-0.2722
0.4	70.7160	72.7809	92.2134	-0.2150
0.6	70.6733	72.5096	116.4531	-0.1702
0.66	71.1241	72.9336	128.2177	-0.1589
0.8	72.7024	74.5151	160.0042	-0.1433
1	76.8066	79	241.2551	-0.1412
1.2	83.2038	86.6224	404.8364	-0.1699
1.4	92.3180	98.7322	768.385	-0.2314
1.6	104.8061	117.9068	1670.897	-0.3205
1.8	121.6161	149.0679	4177.562	-0.4247
2	144.0794	201.7915	11910.4	-0.5288

Al comparar los valores de la estadística T_{λ} con los cuantiles de una normal estándar, es claro que no hay evidencia en contra del modelo logístico. Al igual que en el ejemplo 1, Aguirre, R. (1991) realiza un análisis muy completo de diagnóstico con estos datos, y llega a la conclusión de que no hay ningún tipo de problema con el modelo.

Como se hizo notar en la sección 3.3.2. El uso de la estadística T_λ , cuando hay dispersión extrema, se puede ver como una prueba sobre la función liga propuesta, en este caso la transformación *logit*. En este caso, la estadística T_λ coincide con la estadística basada en puntajes, por lo que si tiene sentido rechazar la hipótesis nula si T_λ es "pequeña". Para el ejemplo que nos ocupa tampoco hay problema en este sentido.

CONCLUSIONES

En este trabajo se hizo una presentación muy general del modelo logístico y algunas de sus propiedades. Este modelo pertenece a una familia muy importante: la familia de los modelos lineales generalizados. Se expusieron algunos conceptos fundamentales sobre estos modelos, y se obtuvieron las expresiones correspondientes del modelo ajustado.

En cuanto al problema de bondad de ajuste, se presentó la estadística $(2N)I^{\lambda}$, cuyo valor depende de un índice λ . Característica importante de esta estadística, que está basada en el concepto de divergencia, es que incluye a otras estadísticas conocidas como caso particular (valores específicos de λ). La obtención de la distribución de $(2N)I^{\lambda}$, bajo la hipótesis nula de que el modelo logístico es el correcto, se complica por la forma misma de la estadística, de tal suerte que hay que recurrir a resultados asintóticos.

Cuando los datos están agrupados, y se cumplen ciertas condiciones límite sobre el tamaño de los grupos, esta última distribución es una $\bar{\chi}^2$. Si los datos no están agrupados, como suele ocurrir al tener covariables continuas, la situación se complica, pues dentro de cada grupo no se puede pensar en tener una distribución χ^2 , al comparar valores observados y esperados. El resultado que se presenta en el Teorema 2, y que fué dado por Rojek en 1969, es importante en esta última situación. La distribución asintótica de $(2N)I^{\lambda}$, cuando el tamaño de la muestra aumenta y en cada grupo hay pocas observaciones, es $N(\mu, \sigma^2)$. Dejando aparte la dificultad que representa la demostración del Teorema, el cálculo de μ , y σ^2 , no es complicado, aunque sí laborioso. Hay condiciones para tener expresiones más sencillas del valor asintótico de la esperanza y la varianza de $(2N)I^{\lambda}$, pero generalmente el comprobar que tales condiciones se cumplan resulta una tarea difícil o imposible. En principio, no se tiene un buen ajuste

cuando el valor de $T_\lambda = ((2N)I^\lambda - \mu_\lambda) / \sigma_\lambda$, es "grande". No obstante, si en cada grupo se tiene una sola observación, la estadística T_λ coincide con la que se obtiene al realizar una prueba sobre la bondad de la función ligo usando la función de puntaje. Esto nos dice que tambien puede tenerse falta de ajuste cuando el valor de T_λ es "pequeño".

Un problema importante en el uso de la estadística $(2N)I^\lambda$, es la selección del valor de λ . Aparte de los casos $\lambda = -2, -1, -0.5, 0, 1$, que dan lugar a estadísticas conocidas, no se tiene todavía un criterio bien definido para elegir el valor de λ . Cressie and Read (1988) sugieren un valor de λ igual a $2/3$, pues se han visto que al incorporar términos de corrección a la estadística $(2N)I^\lambda$, para que sus momentos sean más parecidos a los de una χ^2 , $\lambda = 2/3$ dá lugar a las correcciones más pequeñas.

Para ser consistente con el uso de la estadística $(2N)I^\lambda$, la estimación de los parámetros del modelo se puede realizar minimizando el valor de dicha estadística para algún λ fijo. El estimador que se obtiene es consistente y por lo tanto se puede emplear en el Teorema 2.

El estudiar la potencia de una prueba es sumamente importante, y en este sentido existen resultados parciales que nos aseguran, al menos para alternativas locales, que la distribución asintótica de $(2N)I^\lambda$ es una χ^2 no central para datos agrupados, o bien una normal para datos no agrupados.

Simonoff (1986) sugiere estimar mas eficientemente la varianza de $(2N)I^\lambda$ utilizando el método de jackknife. Este autor trata básicamente el caso de tener un solo grupo formado por las covariables y una respuesta multinomial. Comenta que se espera lo mismo para el caso de tener un modelo lineal generalizado.

Mucha de la información con que se cuenta proviene de haber realizado una encuesta por muestreo. Roberts, Rao y Kumar (1987), se preocupan por tener en cuenta el diseño de muestreo y de acuerdo a esto, ajustar las estadísticas χ^2 y G^2 , al ajustar un modelo logístico. Este trabajo puede ser la base para proponer los ajustes correspondientes a la estadística $(2N)^A$.

BIBLIOGRAFIA

- * Aguirre, R. H. (1991), *Técnicas de Diagnóstico para el Modelo Logístico*. Tesis para obtener el grado de Maestría en Ciencias en el I.I.M.A.S., U.N.A.M.
- * Cook, D.R. (1988), "Assessment of Local Influence". *J. R. Statist. Soc.*, B, 48, 133-169.
- * Cook, D.R. y Weisberg, S. (1982), *Residuals and Influence in Regression*. Chapman and Hall, London.
- * Cox, D.R. (1989) *The Analysis of Binary Data*. Chapman and Hall, London.
- * Cressie, N. y Read, T.R.C. (1984), "Multinomial Goodness-of-fit Test". *J. R. Statist. Soc. B*, 46, 440-464.
- * Chambers, E.A. y Cox, D.R. (1967), "Discrimination between Alternative Binary Response Models". *Biometrika*, 54, 573-8.
- * Chatterjee, S., Hadi, A.S. (1988), *Sensitivity Analysis in Linear Regression*. John Wiley, New York.
- * Dale, J.R. (1986), "Asymptotic Normality of Goodness-of-Fit Statistics for Sparse Product Multinomials". *J. R. Statist. Soc.*, B 48, 48-59.
- * Daniel, C. y Wood, F.S. (1971), *Fitting Equations to Data*. John Wiley, New York.
- * Fahrmeir, L. y Kaufmann, H. (1985), "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models". *The Annals of Statistics*, 13, 342-368.
- * Fisher, R.A. (1924), "The Conditions under which Measures the Discrepancy between Observations and Hypothesis". *J. R. Statist. Soc.* 87, 442-450.
- * Freeman, M.F. y Tukey, J.W. (1950), "Transformations Related to the Angular and Square Root". *Ann. Math. Statist.* 21, 607-611.
- * Guerrero, D.V. y Johnson, R.A. (1982), "Use of the Box-Cox Transformation with Binary Response Models". *Biometrika*, V. 69, 309-314.
- * Hosmer, D. W., y Lemeshow, S. (1989), *Applied Logistic Regression*. John Wiley, New York.

- * Koehler, K.J. (1986), "Goodness-of-Fit for Log-Linear Models in Sparse Contingency Tables". J. Am. Statist. Assoc. 81, 483-493.
- * Kullback, S. (1959), *Information Theory and Statistics*. John Wiley, New York.
- * McCullagh, P. (1985), "On the Asymptotic Distribution of Pearson's Statistics in Linear Exponential Family Models". Int. Statist. Rev. 53, 61-7.
- * McCullagh, P. y Nelder, J.A. (1989), *Generalized Linear Models*. Chapman and Hall, London.
- * Neyman, J. (1959), "Contribution to the Theory of the χ^2 Test". Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability, 239-273.
- * Osius, G. y Rojek, D. (1992), "Normal Goodness-of-Fit Test for Multinomial Models with Large Degrees of Freedom". J. Am. Statist. Assoc. 87, 1145-52.
- * Pearson, K. (1900), "On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to have arisen from Random Sampling". Philosophy Magazine 50, 157-172.
- * Rao, C. R. (1965), *Linear Statistical Inference and its Applications*. John Wiley, New York.
- * Roberts, G., Rao, J.N.K. y Kumar, S. (1987), "Logistic Regression Analysis of Survey Data". Biometrika, V. 74, 1-12.
- * Rojek, D. (1989), *Asymptotik für Anpassungstest in Produkt Multinomial Modellen bei wachsendem Freiheitsgrad*. Tesis Doctoral, Universidad de Bremen, Alemania.
- * Simonoff, J.S. (1986), "Jackknifing and Bootstrapping Goodness-of-Fit Statistics in Sparse Multinomials". J. Am. Statist. Assoc., V. 81, 1005-1011.
- * Searle, S. R. (1971), *Linear Models*. John Wiley, New York.

APENDICE No. 1

Sea Y_i una variable aleatoria cuya distribución pertenece a la familia exponencial, de tal forma que su función de densidad es de la forma

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

Si Y_i es la respuesta de un modelo lineal generalizado, con X matriz de covariables y función līga dada por $\eta = X\beta$, entonces la matriz de información $i(\beta)$ resulta ser

$$i(\beta) = a^{-1}(\phi) X^T (CVC)^{-1} X ,$$

con

$$C = \text{diag} \{ \partial \eta_i / \partial \mu_i \} \quad \text{y} \quad V = \text{diag} \{ b''(\theta_i) \} .$$

La demostración de esta última afirmación se presenta a continuación.

Siendo $L(\beta)$ el logaritmo de la función de verosimilitud y $J(\beta)$ la matriz definida como

$$J(\beta) = \frac{\partial^2 L(\beta)}{\partial \beta_r \partial \beta_s} .$$

el elemento (r,s) de la matriz $i(\beta)$ está dada por

$$\{i(\beta)\}_{rs} = -E\{J(\beta)\}_{rs} = -E \left\{ \frac{\partial^2 L(\beta)}{\partial \beta_r \partial \beta_s} \right\} =$$

$$\begin{aligned}
&= -E \frac{\partial}{\partial \beta_r} \left\{ \sum_{i=1}^n (Y_i - \mu_i) a^{-1}(\phi) \left[b''(\theta_i) \frac{\partial \eta_i}{\partial \mu_i} \right]^{-1} x_{is} \right\} = \\
&= -a^{-1}(\phi) E \sum_{i=1}^n \left[(Y_i - \mu_i) x_{is} \frac{\partial}{\partial \beta_r} \left(b''(\theta_i) \frac{\partial \eta_i}{\partial \mu_i} \right)^{-1} + \left(b''(\theta_i) \frac{\partial \eta_i}{\partial \mu_i} \right) x_{is} \frac{\partial}{\partial \beta_r} (Y_i - \mu_i) \right] = \\
&= a^{-1}(\phi) \left(b''(\theta_i) \frac{\partial \eta_i}{\partial \mu_i} \right)^{-1} x_{is} \frac{\partial \mu_i}{\partial \beta_r} = a^{-1}(\phi) [b''(\theta_i)]^{-1} \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{-2} x_{is} \frac{\partial \eta_i}{\partial \mu_r} \frac{\partial \mu_i}{\partial \beta_r} = \\
&= a^{-1}(\phi) [b''(\theta_i)]^{-1} \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{-2} x_{is} \frac{\partial \eta_i}{\partial \beta_r} = \\
&= a^{-1}(\phi) [b''(\theta_i)]^{-1} \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{-2} x_{is} x_{ir} .
\end{aligned}$$

Por lo que finalmente,

$$I(\beta) = a^{-1}(\phi) X^T (CVC)^{-1} X .$$

APENDICE No. 2

Las condiciones de regularidad a las que se refiere el Teorema 1 fueron establecidas por Birch en 1964. Sea $\phi^* \in \Phi_0$ el verdadero y desconocido valor de ϕ bajo una hipótesis H_0 , y sea $\theta^* \in \Theta_0 \subset \mathfrak{R}^p$, $p < k-1$, tal que $\phi^* = f(\theta^*)$. Las condiciones son,

1. θ^* es un punto interior de Θ_0 .
2. $\phi_i^* = f_i(\theta^*) > 0$, $i = 1, 2, \dots, k$, así que ϕ^* es un punto interior de Φ_0 .
3. El mapeo $f: \Theta_0 \rightarrow \Phi_0$ es totalmente diferenciable en θ^* , de tal suerte que $\partial f_i / \partial \theta_j$ existe en θ^* y además $f(\theta)$ se puede aproximar linealmente de la siguiente forma,

$$f(\theta) = f(\theta^*) + \left[\partial f(\theta^*) / \partial \theta \right] (\theta - \theta^*) + o(\|\theta - \theta^*\|)$$

cuando $\theta \rightarrow \theta^*$.

4. La matriz jacobiana $\partial f(\theta^*) / \partial \theta$ es de rango completo, siendo éste igual a p .
5. El mapeo $f^{-1}: \Phi_0 \rightarrow \Theta_0$ es continuo en $\phi^* = f(\theta^*)$.
6. El mapeo $f: \Theta_0 \rightarrow \Phi_0$ es continuo para todo $\theta^* \in \Theta_0$.

APENDICE No. 3

En términos de la notación empleada en el Capítulo 3 de este trabajo, las condiciones de regularidad de Birch más importantes son:

CR1) El vector de probabilidades π_i , como función del vector β , tiene primera y segunda derivadas continuas con respecto a β .

CR2) La matriz de información $I(\beta)$ es de rango completo.

CR3) $\hat{\beta}$ es consistente, es decir, $\hat{\beta} = \beta + o_p(1)$.

CR4) $\hat{\beta} - \beta = [I(\beta)]^{-1} D I(\beta) + O_p(n^{-1})$. $D I(\beta)$ es la derivada del logaritmo natural de la verosimilitud.

Otra condición que se debe cumplir es sobre la varianza σ_λ^2 , la cual establece que,

CV) $n / \sigma_\lambda^2 < \infty$.

En todo lo sigue de este Apéndice se omite el uso de λ , pues su valor específico no altera los resultados.

La demostración del Teorema 2 comienza con la definición de la suma de desviaciones $Z(\beta) = (2N) I(\beta) - \mu(\beta)$. Una expansión de primer orden dá lugar a

$$Z(\hat{\beta}) - Z(\beta) = D Z(\beta) [\hat{\beta} - \beta] + O_p(1) .$$

En la expresión anterior, la derivada $D Z(\beta)$ se puede reemplazar por su valor esperado, $E\{D Z(\beta)\} = -c(\beta)$, ya que $D Z(\beta) = -c(\beta) + O_p(n^{-1/2})$ (el vector $c(\beta)$ se definió al establecer el Teorema 2). Con esto último y el supuesto CR4, el numerador de la estadística T del Teorema 2 se puede representar como

$$Z(\hat{\beta}) = \psi + O_p(1) , \tag{A3.1}$$

con

$$\psi = Z(\beta) - c(\beta) [i(\beta)]^{-1} Dl(\beta) .$$

Como $Z(\beta)$ y $Dl(\beta)$ son sumas de variables aleatorias independientes, al aplicar el Teorema Central del Límite se obtiene la normalidad asintótica de ψ . Además, $E(\psi) = 0$ y $V(\psi) = \sigma^2(\beta) = v^2(\beta) - Q(\beta)$, así que cuando $n \rightarrow \infty$,

$$\frac{\psi}{\sigma(\beta)} \xrightarrow{d} N(0,1) . \tag{A3.2}$$

A partir de (A3.1) y (A3.2), se tiene lo siguiente,

$$\frac{Z(\hat{\beta})}{\sigma(\beta)} = T \frac{\sigma(\hat{\beta})}{\sigma(\beta)} \xrightarrow{d} N(0,1) ,$$

conforme $n \rightarrow \infty$.

El Teorema 2 quedaría demostrado si se muestra que $\sigma(\hat{\beta})$ es un estimador consistente de $\sigma(\beta)$. Después de obtener algunas propiedades en cuanto al orden de ciertas cantidades, Rojek (1989) demuestra que efectivamente

$$[\sigma(\hat{\beta}) / \sigma(\beta)] \xrightarrow{d} 1 .$$

APENDICE No. 4

En el Capítulo 1 de este trabajo se expresó la estadística $(2N) I^\lambda$ como la suma $\sum_n A_\lambda(Y_i, m_i \pi_i)$, en donde

$$A_\lambda(Y, m\pi) = \frac{2}{\lambda(\lambda+1)} Y \left[\left(\frac{Y}{m\pi} \right)^\lambda - 1 \right] + \frac{2}{\lambda(\lambda+1)} (m-Y) \left[\left(\frac{m-Y}{m(1-\pi)} \right)^\lambda - 1 \right].$$

Sea V_λ la suma

$$V_\lambda = \sum_n [h_\lambda^2(m_i, \pi_i) - c_\lambda^2(m_i, \pi_i) m_i \pi_i^{-1} (1-\pi_i^{-1})],$$

con

$$h_\lambda^2(m, \pi) = V\{A_\lambda(Y, m\pi)\}, \quad c_\lambda^2(m_i, \pi_i) = \text{COV}\{Y, A_\lambda(Y, m\pi)\}.$$

En este Apéndice se demuestra que cuando $m_i = 1$, $i = 1, 2, \dots, n$, V_λ es igual a cero. Primero se calcula $h_\lambda^2(1, \pi)$.

$$\begin{aligned} h_\lambda^2(1, \pi) &= V\{A_\lambda(Y, \pi)\} = E\{A_\lambda^2(Y, m\pi)\} - E^2\{A_\lambda(Y, m\pi)\} = \\ &= E\left\{ \frac{4}{\lambda^2(\lambda+1)^2} Y^2 \left[\left(\frac{Y}{\pi} \right)^\lambda - 1 \right]^2 + \frac{4}{\lambda^2(\lambda+1)^2} (1-Y)^2 \left[\left(\frac{1-Y}{1-\pi} \right)^\lambda - 1 \right]^2 + \right. \\ &\quad \left. + 2 \frac{4}{\lambda^2(\lambda+1)^2} Y(1-Y) \left[\left(\frac{Y}{\pi} \right)^\lambda - 1 \right] \left[\left(\frac{1-Y}{1-\pi} \right)^\lambda - 1 \right] \right\} - \\ &\quad - \left\{ \frac{4}{\lambda^2(\lambda+1)^2} \left[\left(\frac{1}{\pi} \right)^\lambda - 1 \right]^2 Y^2 + \frac{4}{\lambda^2(\lambda+1)^2} \left[\left(\frac{1}{1-\pi} \right)^\lambda - 1 \right]^2 (1-\pi)^2 + \right. \\ &\quad \left. + 2 \frac{4}{\lambda^2(\lambda+1)^2} \left[\left(\frac{1}{\pi} \right)^\lambda - 1 \right] \left[\left(\frac{1}{1-\pi} \right)^\lambda - 1 \right] \pi(1-\pi) \right\} = \end{aligned}$$

$$\begin{aligned}
&= \frac{4}{\lambda^2(\lambda+1)^2} \left[\left(\frac{1}{\pi} \right)^\lambda - 1 \right]^2 \pi(1-\pi) + \frac{4}{\lambda^2(\lambda+1)^2} \left[\left(\frac{1}{1-\pi} \right)^\lambda - 1 \right]^2 (1-\pi)(1-(1-\pi)) - \\
&\quad - 2 \frac{4}{\lambda^2(\lambda+1)^2} \left[\left(\frac{1}{\pi} \right)^\lambda - 1 \right] \left[\left(\frac{1}{1-\pi} \right)^\lambda - 1 \right] \pi(1-\pi) = \\
&= \frac{4\pi(1-\pi)}{\lambda^2(\lambda+1)^2} \left\{ \left[\left(\frac{1}{\pi} \right)^\lambda - \left(\frac{1}{1-\pi} \right)^\lambda \right]^2 \right\}, \tag{A4.1}
\end{aligned}$$

Ahora se calcula $c_\lambda(1, \pi)$.

$$\begin{aligned}
c_\lambda(1, \pi) &= \text{COV}\{Y, A_\lambda(Y, \pi)\} = E\{Y A_\lambda(Y, \pi)\} - E(Y) E\{A_\lambda(Y, \pi)\} = \\
&= \pi A_\lambda(Y, \pi) - \pi \left\{ \frac{2}{\lambda(\lambda+1)} \left[\left(\frac{1}{\pi} \right)^\lambda - 1 \right]^2 \pi + \frac{2}{\lambda(\lambda+1)} \left[\left(\frac{1}{1-\pi} \right)^\lambda - 1 \right]^2 (1-\pi) \right\} = \\
&= \frac{2}{\lambda(\lambda+1)} \left[\left(\frac{1}{\pi} \right)^\lambda - 1 \right]^2 \pi - \frac{2}{\lambda(\lambda+1)} \left[\left(\frac{1}{1-\pi} \right)^\lambda - 1 \right]^2 \pi(1-\pi) = \\
&= \frac{2}{\lambda(\lambda+1)} \pi(1-\pi) \left[\left(\frac{1}{\pi} \right)^\lambda - 1 - \left(\frac{1}{1-\pi} \right)^\lambda + 1 \right] = \\
&= \frac{2}{\lambda(\lambda+1)} \pi(1-\pi) \left[\left(\frac{1}{\pi} \right)^\lambda - \left(\frac{1}{1-\pi} \right)^\lambda \right], \tag{A4.2}
\end{aligned}$$

A partir de (A4.1) y (A4.2) resulta evidente que V_λ es igual a cero, pues cada uno de sus sumandos es igual a cero.