Revista Colombiana de Estadística
Diciembre 2009, volumen 32, no. 2, pp. 157 a 187

Sobre la agrupación de niveles del factor explicativo en el modelo logit binario

About Joining Explanation Factor Levels in the Binary Logit Model

Ernesto Ponsot Balaguer[1,a], Surendra Sinha[1,b], Arnaldo Goitía[1,c]

[1]Programa de Doctorado en Estadística, Instituto de Estadística Aplicada y Computación (IEAC/FACES), Universidad de Los Andes, Mérida, Venezuela

Resumen

Se discute el efecto que se produce sobre el modelo logit binario con un único factor explicativo cuando el investigador decide agrupar algunos niveles de dicho factor. Con base en la parametrización de referencia y el modelo saturado se sugiere un procedimiento que, aprovechando los cómputos de un primer ajuste logit y corrigiendo el supuesto distribucional sobre la varianza, produce estimaciones más eficientemente y con mayor precisión que las que se producen si solo se decide reiterar un ajuste logit. Una vez colocado el tema en perspectiva, se desarrollan las ecuaciones que sustentan el procedimiento sugerido, apelando a la teoría asintótica. Se ilustra mediante un ejemplo la diferencia entre el procedimiento sugerido y el habitual y, con base en una extensa simulación, se muestran tendencias sólidas a favor del primero, en la medida en que las probabilidades de éxito de la variable respuesta ($Y = 1$), asociadas con las categorías del factor explicativo incluidas en la agrupación, sean más disímiles entre sí.

***Palabras clave*:** modelo logit, agregación de niveles, datos agregados, tablas de contingencia, modelo lineal generalizado.

Abstract

We discuss the effect that is produced on the binary logit model with one explanatory factor, when the researcher decides to join some levels of the factor. Based on the reference parametrization and the saturated model a procedure is suggested, that takes advantage of the calculations of the first adjustment and corrects the distribucional supposition around the variance. As a result, it produces estimations more efficiently and with more precision, than those which take place if it is decided to repeat the usual logit fit. Once placed the topic in perspective, we develop the equations that support

[a]Estudiante de doctorado. E-mail: ernesto@ula.ve
[b]Profesor titular. E-mail: sinha32@yahoo.com
[c]Profesor titular. E-mail: goitia@ula.ve

the suggested procedure, based on asymptotic theory. We illustrate with an example the difference between the suggested procedure and the usual one. By developing an extensive simulation, some solid trends appear in favour of the first one, especially when the probabilities of success of the response ($Y = 1$), associated with the categories of the explanatory factor included in the group, are less similar each other.

***Key words*:** Logit model, Joining levels, Aggregate data, Contingency tables, Generalized linear model.

1. Introducción

El modelo logit ha sido en las últimas décadas una herramienta de gran utilidad en el análisis estadístico de datos categóricos y tablas de contingencia (véase por ejemplo Christensen 1997, Powers & Xie 1999, Hosmer & Lemeshow 2000, Agresti 2007, Hilbe 2009). Se desprende como un caso particular del modelo lineal generalizado cuando los datos se suponen distribuidos de forma binomial. Bajo este supuesto, se utiliza logit como función de enlace canónico entre los componentes aleatorio y sistemático del modelo (McCullagh & Nelder 1989, p. 30), y se postulan factores o tratamientos explicativos, al estilo del diseño de experimentos y el análisis de varianza convencionales.

Este modelo propone que el logaritmo de la posibilidad, entendida como el cociente entre la probabilidad de éxito y la de fracaso en un ensayo de Bernoulli, es igual a una función lineal en los parámetros, denominada usualmente predictora lineal. Su propósito es estimar y establecer la significancia estadística de los factores, frente a una respuesta observada. En el proceso, operando con la inversa del logaritmo de posibilidad en función de la predictora lineal, se predicen las probabilidades de éxito en cada combinación de niveles de los factores.

En particular y a efectos de simplificar la exposición, este trabajo supone un modelo logit en su formulación más simple. Esto es, una variable respuesta dicotómica y un solo factor explicativo. Adicionalmente, se asume que las respuestas correspondientes a los distintos niveles del factor explicativo son binomiales independientes.

Es común encontrar en la literatura aplicaciones de este modelo en las más diversas áreas de investigación. Interesa particularmente aquí la situación en que el investigador cuenta con una tabla de contingencia (obtenida en un estudio prospectivo o por muestreo, por ejemplo) y, luego de postular y ajustar un modelo logit a los datos, él mismo decide agrupar algunos niveles del factor y reiterar el análisis, en el sentido de ajustar nuevamente un modelo logit sobre la tabla de contingencia resultante de la agrupación.

Por ejemplo, este procedimiento es sugerido en Hosmer & Lemeshow (2000, p. 136) como estrategia para subsanar el inconveniente de respuestas con muy baja o ninguna representación en la tabla de contingencia. Esta eventualidad ocasiona problemas numéricos ya que, cuando su ajuste se lleva a cabo por medios asintó-

ticos, la estimación de los parámetros del modelo logit es exigente con respecto a los tamaños de muestra.

También abundan los ejemplos en que el investigador agrega niveles del factor, simplemente para disminuir la complejidad del análisis o bien por cuanto le interesa a posteriori concentrarse en algunos niveles y tratar los restantes de forma anónima. Un ejercicio que ilustra este proceder puede verse en Hilbe (2009, pp. 74 y 88). En su texto, de origen muy reciente, el autor desarrolla modelos a partir del Registro Nacional Cardiovascular del Canadá, empleando en una primera oportunidad la edad con cuatro niveles como factor explicativo, y en otra oportunidad agrupando dicho factor hasta alcanzar solo dos niveles.

Al reiterar el ajuste de un modelo logit sobre una segunda tabla de contingencia con niveles agrupados de los factores, en general se incurre en una violación del supuesto binomial original, con implicaciones importantes sobre las varianzas estimadas. Este es el centro de la investigación pautada en la tesis doctoral del primer autor (Ponsot 2009), uno de cuyos resultados parciales es el presente trabajo.

Consecuentemente, el propósito de este artículo es ahondar sobre dicho problema y, manteniéndose en el ámbito del modelo logit, sugerir un nuevo curso de acción que mejore la precisión de los resultados. La exposición continúa con la siguiente sección dedicada a la formulación del problema. La tercera sección se destina al estudio de las varianzas, tanto cuando se supone que todas las poblaciones son binomiales (el procedimiento habitual), como cuando no es así (el procedimiento sugerido). En la cuarta sección se presenta el procedimiento sugerido cuando el investigador agrupa los dos últimos niveles, argumentando con base en la teoría asintótica. En la quinta sección se comenta la extensión natural del procedimiento sugerido, cuando se agrupan en general k niveles ($k \geq 2$). La sexta sección del trabajo contiene una comparación entre ambos procedimientos, ilustrada mediante un ejemplo concreto. La séptima sección sintetiza los resultados de una simulación extensa sobre 10000 tablas de contingencia generadas seudoaleatoriamente, cuyo código fuente R se reproduce en el anexo. Por último, la octava sección se dedica a las conclusiones.

2. Formulación del problema

Tabla 1: $Y(0,1)$ *vs.* $A(1,\ldots,a)$.

	Y		
A	0	1	Total
1	n_{10}	n_{11}	$n_{1\cdot}$
2	n_{20}	n_{21}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
a	n_{a0}	n_{a1}	$n_{a\cdot}$
Total	$n_{\cdot 0}$	$n_{\cdot 1}$	$n_{\cdot\cdot}$

Sea la tabla 1 un arreglo de la variable categórica Y (respuesta binaria) versus los niveles del factor A (nominales u ordinales), en el cual n_{ij} ($i = 1,\ldots,a$ y

$j = 0, 1$) representa la frecuencia simple de aparición del i-ésimo nivel del factor A y la j-ésima categoría de la variable respuesta Y, entendiendo $Y = 0$ como fracaso y $Y = 1$ como éxito.

En presencia de una respuesta binaria, la tabla 1 queda completamente especificada con los valores n_{i1} y $n_{i\cdot}$, ya que $n_{i\cdot} = n_{i0} + n_{i1}$. Para simplificar la notación, sea entonces $y_i \equiv n_{i1}$ el número de éxitos observado en el i-ésimo nivel de A, y $t_i \equiv n_{i\cdot}$ el total de observaciones para dicho nivel.

La situación de interés en este trabajo asume que las respuestas correspondientes a los distintos niveles de A son independientes entre sí. También se supone que dichas respuestas, es decir, las frecuencias observadas de los niveles del factor, provienen de una población binomial en el número de éxitos ($Y = 1$), esto es,

$$Y_i \overset{Ind}{\sim} \text{Bin}(t_i, p_i),\ \forall\ i = 1, \ldots, a$$

donde Y_i es la variable aleatoria que representa el número de éxitos en la i-ésima muestra y p_i, considerada constante, es la probabilidad de éxito asociada. Se supondrá además en este trabajo la inexistencia de sobredispersión.

Entre muchos modelos que pueden formularse para el caso, es de interés particular aquí el modelo logit. Dicho modelo es la versión del tipo análisis de varianza (ANOVA) del modelo de regresión logística, y aun cuando se conoce desde hace varias décadas, en la actualidad (Hilbe 2009, p. 4) se suele presentar como un caso particular del modelo lineal generalizado, originalmente propuesto por Nelder & Wedderburn (1972). Su formulación es la siguiente:

$$\text{logit}(p_i) = \boldsymbol{x_i'}\boldsymbol{\beta},\ \ i = 1, \ldots, a \tag{1}$$

En la ecuación (1), $\boldsymbol{\beta} = \begin{bmatrix}\beta_1 & \beta_2 & \cdots & \beta_r\end{bmatrix}'$ es un vector columna de parámetros desconocidos, con r un número entero fijo que representa la cantidad de dichos parámetros que el investigador ha decidido incorporar en su diseño; $\boldsymbol{x_i'}$ es el i-ésimo vector fila de la matriz de diseño $X_{a\times r}$ la cual, en presencia del modelo saturado, como es el caso, resulta en una matriz cuadrada ($r = a$).

La expresión $\text{logit}(p_i)$ es la aplicación de la transformación $\text{logit}(p) = \log_e[p/(1-p)]$ a las probabilidades de éxito poblacionales supuestas para la i-ésima muestra. Para simplificar la notación del modelo en términos matriciales y hacer compatibles los subíndices de los elementos de la matriz X y el vector $\boldsymbol{\beta}$, el parámetro β_1 se corresponde en esta formulación con el intercepto (usualmente denotado por β_0). Las cantidades $p_i/(1 - p_i)$ se interpretan como las "posibilidades" (en inglés, *odds*) del éxito frente al fracaso. Consecuentemente $\text{logit}(p_i)$ modela el logaritmo neperiano de la posibilidad en la i-ésima muestra. Claramente las p_i son también objeto de estimación, por lo cual el ajuste del modelo se produce generalmente por la aplicación del método de Newton-Raphson, entre otros, que implica un cómputo iterativo hasta lograr la convergencia. Otro procedimiento de estimación, no iterativo, se conoce en la literatura como el método GSK, originalmente propuesto por Grizzle et al. (1969).

En su acepción más simple es común utilizar la parametrización de referencia para la matriz X. Sin pérdida de generalidad, sea a el nivel de referencia, entonces

dicha parametrización conduce al modelo siguiente:

$$\begin{bmatrix} \text{logit}(p_1) \\ \text{logit}(p_2) \\ \vdots \\ \text{logit}(p_{a-1}) \\ \text{logit}(p_a) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{a-1} \\ \beta_a \end{bmatrix} = \begin{bmatrix} \beta_1 + \beta_2 \\ \beta_1 + \beta_3 \\ \vdots \\ \beta_1 + \beta_a \\ \beta_1 \end{bmatrix} \tag{2}$$

En (2), el modelo es saturado ($a = r$) y por lo tanto no se cuenta con los grados de libertad suficientes para que tenga sentido el cálculo de los estadísticos de devianza de Pearson. Sin embargo, aún pueden estimarse sus parámetros ($\boldsymbol{\beta}$) y determinarse su significación estadística. Nótese además que existe X^{-1} puesto que X es no singular.

En efecto, el cálculo del determinante por la descomposición en cofactores (X_{ij}), por simplicidad, pivotando la última fila de la matriz X ya que el único de sus elementos diferente de cero es x_{a1}, resulta $|X| = (-1)^{a+1}|I| = (-1)^{a+1} \neq 0$.

El modelo logit ha sido profusamente estudiado y caracterizado (véanse por ejemplo Cox 1970, McCullagh & Nelder 1989, McCulloch & Searle 2001, Collett 2002, Agresti 2007, Rodríguez 2008). Sin embargo, poco se ha profundizado en las consecuencias que acarrea la violación de sus supuestos. En particular, son del interés aquí las consecuencias de la violación del supuesto binomial para las poblaciones subyacentes.

Así, supóngase que el investigador decide agrupar los niveles a y $a-1$, haciendo $y^*_{a-1} = y_{a-1} + y_a$ y $t^*_{a-1} = t_{a-1} + t_a$. Claramente, la situación puede extenderse a más de dos niveles, simplemente agregando los dos últimos, luego estos con el anterior y así sucesivamente. El teorema 1 establece que la suma de dos variables aleatorias independientes binomiales, con probabilidades de éxito en general distintas, no resulta en una variable aleatoria binomial.

Teorema 1. *Sean X_1 y X_2 dos variables aleatorias independientes tales que $X_1 \sim Bin(n_1, p_1)$ y $X_2 \sim Bin(n_2, p_2)$ con $n_1 \leq n_2$. Entonces, la variable aleatoria $Z = X_1 + X_2$ de distribuye como sigue:*

$$P[Z = k] = \left(\frac{p_1}{1 - p_1}\right)^k (1 - p_1)^{n_1} (1 - p_2)^{n_2} S(k) \tag{3}$$

donde

$$S(k) = \begin{cases} \displaystyle\sum_{i=0}^{k} \binom{n_1}{k-i} \binom{n_2}{i} \left[\frac{p_2(1-p_1)}{p_1(1-p_2)}\right]^i, k = 0, \ldots, n_1 \\ \displaystyle\sum_{i=k-n_1}^{k} \binom{n_1}{k-i} \binom{n_2}{i} \left[\frac{p_2(1-p_1)}{p_1(1-p_2)}\right]^i, k = n_1 + 1, \ldots, n_2 \\ \displaystyle\sum_{i=k-n_1}^{n_2} \binom{n_1}{k-i} \binom{n_2}{i} \left[\frac{p_2(1-p_1)}{p_1(1-p_2)}\right]^i, k = n_2 + 1, \ldots, n_1 + n_2 \end{cases}$$

Demostración.

$$
\begin{aligned}
\mathrm{P}[Z=0] &= \mathrm{P}[X_1=0, X_2=0] = \sum_{i=0}^{0} \mathrm{P}[X_1=0-i, X_2=i] \\
\mathrm{P}[Z=1] &= \mathrm{P}[X_1=1, X_2=0] + \mathrm{P}[X_1=0, X_2=1] \\
&= \sum_{i=0}^{1} \mathrm{P}[X_1=1-i, X_2=i] \\
&\vdots \\
\mathrm{P}[Z=n_1] &= \mathrm{P}[X_1=n_1, X_2=0] + \cdots + \mathrm{P}[X_1=0, X_2=n_1] \\
&= \sum_{i=0}^{n_1} \mathrm{P}[X_1=n_1-i, X_2=i] \\
\mathrm{P}[Z=n_1+1] &= \mathrm{P}[X_1=n_1, X_2=1] + \cdots + \mathrm{P}[X_1=0, X_2=n_1+1] \\
&= \sum_{i=1}^{n_1+1} \mathrm{P}[X_1=n_1+1-i, X_2=i] \\
&\vdots \\
\mathrm{P}[Z=n_2] &= \mathrm{P}[X_1=n_1, X_2=n_2-n_1] + \cdots + \mathrm{P}[X_1=0, X_2=n_2] \\
&= \sum_{i=n_2-n_1}^{n_2} \mathrm{P}[X_1=n_2-i, X_2=i] \\
\mathrm{P}[Z=n_2+1] &= \mathrm{P}[X_1=n_1, X_2=n_2-n_1+1] + \cdots + \mathrm{P}[X_1=1, X_2=n_2] \\
&= \sum_{i=n_2-n_1+1}^{n_2} \mathrm{P}[X_1=n_2+1-i, X_2=i] \\
&\vdots \\
\mathrm{P}[Z=n_2+n_1] &= \mathrm{P}[X_1=n_1, X_2=n_2] = \sum_{i=n_2}^{n_2} \mathrm{P}[X_1=n_1+n_2-i, X_2=i]
\end{aligned}
$$

$$
\therefore\ \mathrm{P}[Z=k] \begin{cases} \sum_{i=0}^{k} \mathrm{P}[X_1=k-i, X_2=i], & k=0,\ldots,n_1 \\ \sum_{i=k-n_1}^{k} \mathrm{P}[X_1=k-i, X_2=i], & k=n_1+1,\ldots,n_2 \\ \sum_{i=k-n_1}^{n_2} \mathrm{P}[X_1=k-i, X_2=i], & k=n_2+1,\ldots,n_1+n_2 \end{cases}
$$

Ahora bien, como X_1 y X_2 son independientes, para $r = 0, 1, \ldots, n_1$ y $s = 0, 1, \ldots, n_2$, se tiene que:

$$
\mathrm{P}[X_1=r, X_2=s] = \binom{n_1}{r} p_1^r (1-p_1)^{n_1-r} \binom{n_2}{s} p_2^s (1-p_2)^{n_2-s}
$$

Luego, para un k fijo, llámese $i(k)$ al rango correspondiente de cada sumatoria implicada en $\mathrm{P}[Z=k]$. Entonces:

$$\begin{aligned}\sum_{i(k)} \mathrm{P}[X_1 = k-i, X_2 = i] &= \sum_{i(k)} \binom{n_1}{k-i} p_1^{k-i}(1-p_1)^{n_1-k+i} \binom{n_2}{i} p_2^i (1-p_2)^{n_2-i} \\ &= \left(\frac{p_1}{1-p_1}\right)^k (1-p_1)^{n_1}(1-p_2)^{n_2} \\ &\qquad \sum_{i(k)} \binom{n_1}{k-i}\binom{n_2}{i}\left[\frac{p_2(1-p_1)}{p_1(1-p_2)}\right]^i\end{aligned}$$

Consecuentemente,

$$\mathrm{P}[Z=k] = \left(\frac{p_1}{1-p_1}\right)^k (1-p_1)^{n_1}(1-p_2)^{n_2} S(k), \text{ donde:}$$

$$S(k) = \begin{cases} \sum_{i=0}^{k} \binom{n_1}{k-i}\binom{n_2}{i}\left[\frac{p_2(1-p_1)}{p_1(1-p_2)}\right]^i, k = 0,\ldots,n_1 \\ \sum_{i=k-n_1}^{k} \binom{n_1}{k-i}\binom{n_2}{i}\left[\frac{p_2(1-p_1)}{p_1(1-p_2)}\right]^i, k = n_1+1,\ldots,n_2 \\ \sum_{i=k-n_1}^{n_2} \binom{n_1}{k-i}\binom{n_2}{i}\left[\frac{p_2(1-p_1)}{p_1(1-p_2)}\right]^i, k = n_2+1,\ldots,n_1+n_2 \end{cases}$$

□

El corolario 1 establece, empleando el teorema 1, que la distribución binomial se obtiene cuando las probabilidades de éxito son iguales.

Corolario 1. $p_1 = p_2 = p \Rightarrow Z \sim Bin(n = n_1 + n_2, p)$.

Demostración.

$$\begin{aligned}\mathrm{P}[Z=k] &= \left(\frac{p_1}{1-p_1}\right)^k (1-p_1)^{n_1}(1-p_2)^{n_2} S(k) \\ &= p^k (1-p)^{n-k} S(k)\end{aligned}$$

Y ya que $\{[p_2(1-p_1)]/[p_1(1-p_2)]\}^i = \{[p(1-p)]/[p(1-p)]\}^i = 1$ para i finito,

$$S(k) = \begin{cases} \sum_{i=0}^{k} \binom{n_1}{k-i}\binom{n_2}{i}, k = 0,\ldots,n_1 \\ \sum_{i=k-n_1}^{k} \binom{n_1}{k-i}\binom{n_2}{i}, k = n_1+1,\ldots,n_2 \\ \sum_{i=k-n_1}^{n_2} \binom{n_1}{k-i}\binom{n_2}{i}, k = n_2+1,\ldots,n_1+n_2 \end{cases}$$

Ahora, haciendo uso de propiedades combinatorias (Rohatgi & Ehsanes 2001, Feller 1968), se tiene:

a) Para $k = 0, \ldots, n_1$

$$S(k) = \sum_{i=0}^{k} \binom{n_1}{k-i}\binom{n_2}{i} = \binom{n_1+n_2}{k} = \binom{n}{k}$$

b) Para $k = n_1 + 1, \ldots, n_2$

$$\begin{aligned} S(k) &= \sum_{i=k-n_1}^{k} \binom{n_1}{k-i}\binom{n_2}{i} \\ &= \binom{n_1+n_2-n_2}{n_1}\binom{n_2}{k-n_1} + \binom{n_1+n_2-n_2}{n_1-1}\binom{n_2}{k-(n_1-1)} + \cdots \\ &+ \binom{n_1+n_2-n_2}{0}\binom{n_2}{k} = \binom{n_1+n_2}{k} = \binom{n}{k} \end{aligned}$$

c) Para $k = n_2 + 1, \ldots, n_1 + n_2$

$$\begin{aligned} S(k) &= \sum_{i=k-n_1}^{n_2} \binom{n_1}{k-i}\binom{n_2}{i} \\ &= \binom{n_1+n_2-n_2}{n_1}\binom{n_2}{k-n_1} + \binom{n_1+n_2-n_2}{n_1-1}\binom{n_2}{k-(n_1-1)} + \cdots \\ &+ \binom{n_1+n_2-n_2}{k-n_2}\binom{n_2}{n_2} = \binom{n_1+n_2}{k} = \binom{n}{k} \end{aligned}$$

$$\therefore P[Z = k] = p^k(1-p)^{n-k}S(k) = \binom{n}{k}p^k(1-p)^{n-k} \qquad \square$$

Corolario 2. *$E[Z] = n_1p_1 + n_2p_2$ y $V[Z] = n_1p_1(1-p_1) + n_2p_2(1-p_2)$.*

Demostración. Por una parte,

$$\mathrm{E}[Z] = \mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2] = n_1p_1 + n_2p_2$$

y dado que X_1 y X_2 son independientes, entonces:

$$\mathrm{V}[Z] = \mathrm{V}[X_1 + X_2] = \mathrm{V}[X_1] + \mathrm{V}[X_2] = n_1p_1(1-p_1) + n_2p_2(1-p_2) \qquad \square$$

La ecuación (3) es una generalización de la suma de ensayos de Poisson (Feller 1968, p. 218), es decir, ensayos independientes de Bernoulli con probabilidades de éxito en general diferentes. Se trata de una realización particular de la distribución de probabilidades conocida en la literatura como Poisson-Binomial (Wang 1993, p. 298), y ha sido estudiada desde varios puntos de vista a partir de la aparición de los

trabajos de Neyman (1939). No obstante, al no tener una forma analítica simple, ha recibido atención casi exclusivamente por la vía de las aproximaciones numéricas (Sprott 1958, Hodges & Le Cam 1960, Ollero & Ramos 1991, Weba 1999, Roos 1999, Neammanee 2005). En particular, la estimación de sus parámetros no presenta problemas desde el punto de vista numérico, menos aún hoy en día con la gran capacidad computacional disponible. Sin embargo, su tratamiento analítico, por ejemplo, como factor en la función de verosimilitud, es de complejidad considerable.

Claramente, establecidos los supuestos de un primer modelo logit sobre la tabla 1, un segundo modelo logit sobre una nueva tabla que agrupe los niveles $a-1$ y a, en el caso en que $p_{a-1} \neq p_a$, violenta el supuesto binomial original en la muestra correspondiente al último de los niveles de la variable respuesta, es decir, aquel que surge de la agrupación. Además, puede descartarse su pertenencia a la familia exponencial, ya que la conformación de la densidad de probabilidades depende del recorrido de la variable aleatoria como se aprecia en (3).

Este obstáculo es de consideración puesto que el modelo logit, en el contexto del modelo lineal generalizado, supone una función de enlace canónico deducida a partir de la pertenencia de la distribución de la muestra a la familia exponencial. Una alternativa sería entonces proceder con base en la función de cuasiverosimilitud (Wedderburn 1974) en lugar de la función de verosimilitud. Sin embargo, tampoco es posible encontrar una relación funcional entre la media y la varianza exclusivamente para la distribución en (3), con lo cual queda descartada esta posibilidad (McCullagh & Nelder 1989, p. 337).

Procurando mantener el problema en el ámbito del modelo lineal generalizado y modelo logit, este trabajo persigue entonces estudiar las dos aristas siguientes:

1. Cuál es el efecto de la agregación de niveles sobre las estimaciones y las pruebas de hipótesis y, para el caso en que tal efecto resulte importante, cómo pueden mejorarse los resultados en el sentido de la disminución en la longitud de los intervalos de confianza estimados o, equivalentemente, aumentando la precisión de los estimadores.

2. Cómo puede aprovecharse la información obtenida a partir de un primer modelo logit en el ajuste de un segundo modelo con niveles agregados del factor.

Para ello se adoptan argumentos de naturaleza asintótica, basados en el teorema del Límite Central (TLC) (Lehmann 1999, p. 73) y el método delta (Lehmann 1999, p. 86).

3. Estudio de las varianzas

Como se dijo, el modelo logit supone una distribución binomial en el número de éxitos de la variable respuesta en cada nivel del factor explicativo. Ello implica la suposición $\mathrm{V}[Y_i] = t_i p_i(1-p_i)$ para todo $i = 1, \ldots, a$. Sin embargo, cuando el

investigador agrupa dos niveles cualesquiera, por ejemplo $a-1$ y a, formando una nueva variable aleatoria $Y^*_{a-1} = Y_{a-1} + Y_a$, y nuevamente ejecuta el procedimiento de ajuste del modelo logit, implícitamente supone la varianza como $\mathrm{V}_{\mathrm{Bin}}[Y^*_{a-1}] = t^*_{a-1}p^*_{a-1}(1-p^*_{a-1})$, donde $t^*_{a-1} = t_{a-1} + t_a$ y $p^*_{a-1} = \mathrm{E}[Y^*_{a-1}]/t^*_{a-1} = (t_{a-1}p_{a-1} + t_a p_a)/(t_{a-1} + t_a)$.

Ahora bien, como se prueba en el corolario 2, una expresión adecuada para la varianza es $\mathrm{V}[Y^*_{a-1}] = t_{a-1}p_{a-1}(1-p_{a-1}) + t_a p_a(1-p_a)$. El teorema 2 muestra que ambas expresiones de la varianza no son equivalentes y de hecho, para valores dados de los parámetros, se tiene que $\mathrm{V}_{\mathrm{Bin}}[Y^*_{a-1}] \geq \mathrm{V}[Y^*_{a-1}]$.

Teorema 2. *Con $Y^*_{a-1} = Y_{a-1} + Y_a$, para valores dados de $t_{a-1}, t_a, p_{a-1}, p_a$, $V_{Bin}[Y^*_{a-1}] \geq V[Y^*_{a-1}]$.*

Demostración.

$$\begin{aligned}
\mathrm{V}_{\mathrm{Bin}}[Y^*_{a-1}] &= t^*_{a-1}p^*_{a-1}(1-p^*_{a-1}) \\
&= (t_{a-1}+t_a)\left(\frac{t_{a-1}p_{a-1}+t_a p_a}{t_{a-1}+t_a}\right)\left(1-\frac{t_{a-1}p_{a-1}+t_a p_a}{t_{a-1}+t_a}\right) \\
&= \frac{t^2_{a-1}p_{a-1}(1-p_{a-1}) + t^2_a p_a(1-p_a) + t_{a-1}t_a[p_{a-1}+p_a-2p_{a-1}p_a]}{t_{a-1}+t_a} \\
&= \frac{t_{a-1}\mathrm{V}[Y_{a-1}] + t_a\mathrm{V}[Y_a] + t_{a-1}t_a[p_{a-1}+p_a-2p_{a-1}p_a]}{t_{a-1}+t_a}
\end{aligned}$$

Sea ΔV el incremento en varianza entre ambos supuestos, definido como $\Delta V = \mathrm{V}_{\mathrm{Bin}}[Y^*_{a-1}] - \mathrm{V}[Y^*_{a-1}]$. Entonces:

$$\begin{aligned}
\Delta V &= \frac{t_{a-1}\mathrm{V}[Y_{a-1}] + t_a\mathrm{V}[Y_a] + t_{a-1}t_a[p_{a-1}+p_a-2p_{a-1}p_a]}{t_{a-1}+t_a} - \mathrm{V}[Y_{a-1}] - \mathrm{V}[Y_a] \\
&= \frac{t_{a-1}t_a[p_{a-1}+p_a-2p_{a-1}p_a] - t_a\mathrm{V}[Y_{a-1}] - t_{a-1}\mathrm{V}[Y_a]}{t_{a-1}+t_a} \\
&= \frac{t_{a-1}t_a[p_{a-1}+p_a-2p_{a-1}p_a] - t_a t_{a-1}p_{a-1} + t_a t_{a-1}p^2_{a-1} - t_{a-1}t_a p_a + t_{a-1}t_a p^2_a}{t_{a-1}+t_a} \\
&= \frac{-2t_{a-1}t_a p_{a-1}p_a + t_a t_{a-1}p^2_{a-1} + t_{a-1}t_a p^2_a}{t_{a-1}+t_a} \\
&= \frac{t_{a-1}t_a}{t_{a-1}+t_a}(p_{a-1}-p_a)^2 \qquad (4)
\end{aligned}$$

Claramente $\Delta V \geq 0 \Rightarrow \mathrm{V}_{\mathrm{Bin}}[Y^*_{a-1}] \geq \mathrm{V}[Y^*_{a-1}]$. En particular, si $p_{a-1} = p_a \Rightarrow \mathrm{V}_{\mathrm{Bin}}[Y^*_{a-1}] = \mathrm{V}[Y^*_{a-1}]$. □

La figura 1 ilustra el comportamiento de ambas varianzas cuando se fijan los parámetros t_{a-1}, t_a, p_{a-1}, evaluando la cantidad p_a en los niveles $0.3, 0.5, 0.8$.

Una discusión interesante sobre el tema, partiendo de una sucesión de ensayos de Poisson, puede consultarse en Nedelman & Wallenius (1986) y en las referencias citadas allí. Ahora bien, del examen de la ecuación (4) se desprenden dos hechos importantes:

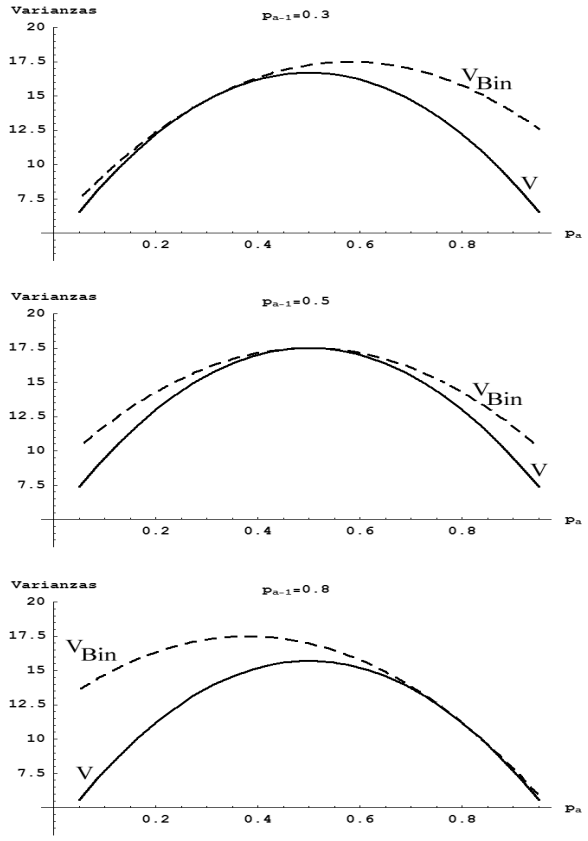

FIGURA 1: Comparación de $\mathrm{V}_{\mathrm{Bin}}[Y^*_{a-1}]$ y $\mathrm{V}[Y^*_{a-1}]$ con $t_{a-1} = 20$, $t_a = 50$.

1. ΔV es directamente proporcional al cuadrado de la diferencia entre p_{a-1} y p_a, siempre positivo o nulo (llámese $\Delta p = |p_{a-1} - p_a|$). Por consiguiente, sin importar cuáles sean los valores de p_{a-1} y p_a que originan Δp, mientras más distantes estén entre sí, tanto mayor será la diferencia entre ambas varianzas.

2. ΔV es también directamente proporcional a la función:

$$T(t_{a-1}, t_a) = t_{a-1}t_a/(t_{a-1} + t_a)$$

estrictamente positiva y creciente. Por lo tanto, la diferencia entre ambas varianzas se incrementará en la medida en que aumenten simultáneamente t_{a-1} y t_a, o cualquiera de ellos, manteniéndose constante el otro. Además, sin pérdida de generalidad, supóngase que $t_{a-1} \leq t_a$. En tal caso T es máxima cuando $t_{a-1} = t_a = t$ y $\Delta V = t(\Delta p)^2/2$.

Así, es clara la existencia de un problema cuando se agrupan niveles y se insiste en el modelo logit sin variar el supuesto binomial, especialmente en lo relacionado con la estimación de las varianzas y, en consecuencia, de los errores estándares utilizados para las pruebas de hipótesis sobre los parámetros del modelo.

4. El procedimiento sugerido

La estimación máximo-verosímil de los parámetros en el modelo logit (1) se lleva a cabo, bajo el supuesto binomial, mediante por ejemplo la aplicación del método de Newton-Raphson al sistema de ecuaciones de las primeras derivadas parciales del logaritmo de la función de verosimilitud, respecto a los β_j $(j = 1, \ldots, r)$, igualadas a cero. Las probabilidades de éxito poblacionales en la verosimilitud se sustituyen, despejando del modelo, como $p_i = e^{x_i'\beta}/(1+e^{x_i'\beta})$ para $i = 1, \ldots, a$. En el procedimiento se itera hasta tanto la diferencia entre dos estimaciones sucesivas de $\boldsymbol{\beta}$ sea despreciable.

La última iteración del método produce las estimaciones máximo-verosímiles tanto de p_i como de $\boldsymbol{\beta}$, cuyos estimadores se denotan respectivamente por $\widehat{p}_i$ y $\widehat{\boldsymbol{\beta}}$. McCullagh & Nelder (1989, p. 119) demuestran que, asintóticamente, es decir, cuando los y_i son suficientemente grandes, el vector $\widehat{\boldsymbol{\beta}}$ obtenido por este mecanismo es un estimador insesgado de $\boldsymbol{\beta}$ y $\mathrm{V}\Big[\widehat{\boldsymbol{\beta}}\Big] = (X'WX)^{-1}$, donde $W = \mathrm{diag}\big[t_i p_i(1 - p_i)\big]$. Este resultado es válido para cualesquiera matriz de diseño $X_{a\times r}$ y vector columna de parámetros $\boldsymbol{\beta}_r$. En particular, es válido también para el caso del modelo saturado en el cual $a = r$.

Luego, en presencia del modelo saturado (2) que postula la equivalencia entre $\mathrm{logit}(p_i)$ y la función predictora lineal $\boldsymbol{x_i'\beta}$ (en términos matriciales $X\boldsymbol{\beta}$), sustituyendo los parámetros por sus estimadores se tiene que, asintóticamente, $\mathrm{V}\Big[X\widehat{\boldsymbol{\beta}}\Big] = X(X'WX)^{-1}X' = XX^{-1}W^{-1}(X')^{-1}X' = W^{-1}$. Por consiguiente y como consecuencia del TLC, así como de las propiedades de los estimadores máximo-verosímiles, se tiene que:

$$\mathrm{logit}(\widehat{p}_i) = \boldsymbol{x_i'}\widehat{\boldsymbol{\beta}} \sim \mathrm{AN}\Big(\boldsymbol{x_i'\beta},\, \big[t_i p_i(1 - p_i)\big]^{-1}\Big) \tag{5}$$

Nótese que tanto en la ecuación (5) como en las que siguen se ha decidido emplear la notación "AN", frecuente en la literatura estadística, como abreviatura de la frase "asintóticamente normal". Ahora bien, empleando el método delta a partir de la ecuación (5) se tiene:

$$\begin{aligned}
\widehat{p}_i &= \frac{e^{\boldsymbol{x_i'}\widehat{\boldsymbol{\beta}}}}{1 + e^{\boldsymbol{x_i'}\widehat{\boldsymbol{\beta}}}} = g^{-1}\Big(\boldsymbol{x_i'}\widehat{\boldsymbol{\beta}}\Big) \\
\frac{\mathrm{d}\, g^{-1}(\boldsymbol{x_i'\beta})}{\mathrm{d}\,(\boldsymbol{x_i'\beta})} &= \frac{e^{\boldsymbol{x_i'\beta}}\big(1 + e^{\boldsymbol{x_i'\beta}}\big) - e^{\boldsymbol{x_i'\beta}}e^{\boldsymbol{x_i'\beta}}}{\big(1 + e^{\boldsymbol{x_i'\beta}}\big)^2} = \frac{e^{\boldsymbol{x_i'\beta}}}{\big(1 + e^{\boldsymbol{x_i'\beta}}\big)^2} = p_i(1 - p_i) \\
\Rightarrow \widehat{p}_i &\sim \mathrm{AN}\left(p_i = \frac{e^{\boldsymbol{x_i'\beta}}}{1 + e^{\boldsymbol{x_i'\beta}}};\, \frac{\big[p_i(1 - p_i)\big]^2}{t_i p_i(1 - p_i)} = \frac{p_i(1 - p_i)}{t_i}\right)
\end{aligned} \tag{6}$$

ya que la función $g^{-1}(\cdot)$ es no nula y diferenciable.

Entonces, la ecuación (6) implica que el investigador conoce la distribución asintótica del estimador de las probabilidades de éxito para el modelo (2). En

particular, cumplido el procedimiento de ajuste, el investigador tiene la estimación mediante el modelo logit de p_i, $\widehat{p}_i = e^{\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}} \Big/ \left[1 + e^{\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}}\right]$ y la varianza asintótica del estimador $V[\widehat{p}_i] = p_i(1-p_i)/t_i$.

Ahora bien, siguiendo la notación definida en la sección anterior, cuando se agrupan los niveles $a-1$ y a, el estimador máximo-verosímil de la nueva probabilidad de éxito p^*_{a-1} es

$$\widehat{p^*}_{a-1} = \frac{\widehat{\mathrm{E}[Y^*_{a-1}]}}{t^*_{a-1}} = \frac{\widehat{\mathrm{E}[Y_{a-1}]} + \widehat{\mathrm{E}[Y_a]}}{t_{a-1}+t_a} = \frac{t_{a-1}\widehat{p}_{a-1} + t_a\widehat{p}_a}{t_{a-1}+t_a} \tag{7}$$

Considerando el corolario 2, claramente $\widehat{p^*}_{a-1}$ es un estimador insesgado de p^*_{a-1}. El teorema 3 demuestra cómo se distribuye este nuevo estimador cuando t_{a-1} y t_a son suficientemente grandes.

Teorema 3. *Si los $\widehat{p}_i$ se distribuyen independientes como en* (6) *para $i = a-1$ e $i = a$, entonces:*

$$\begin{aligned}\widehat{p^*}_{a-1} &= \frac{t_{a-1}\widehat{p}_{a-1} + t_a\widehat{p}_a}{t_{a-1}+t_a} \\ &\sim AN\left(\frac{t_{a-1}p_{a-1} + t_ap_a}{t_{a-1}+t_a}; \frac{t_{a-1}p_{a-1}(1-p_{a-1}) + t_ap_a(1-p_a)}{(t_{a-1}+t_a)^2}\right)\end{aligned}$$

Demostración. Claramente, en el límite, $\widehat{p^*}_{a-1}$ es la suma ponderada de dos funciones lineales de variables aleatorias independientes asintóticamente normales. Luego, su distribución asintótica también es normal. Ahora bien, por un lado se tiene que:

$$\mathrm{E}\left[\widehat{p^*}_{a-1}\right] = \frac{t_{a-1}\mathrm{E}[\widehat{p}_{a-1}] + t_a\mathrm{E}[\widehat{p}_a]}{t_{a-1}+t_a} = \frac{t_{a-1}p_{a-1} + t_ap_a}{t_{a-1}+t_a}$$

y por el otro:

$$\mathrm{V}\left[\widehat{p^*}_{a-1}\right] = \frac{t^2_{a-1}\mathrm{V}[\widehat{p}_{a-1}] + t^2_a\mathrm{V}[\widehat{p}_a]}{(t_{a-1}+t_a)^2} = \frac{t_{a-1}p_{a-1}(1-p_{a-1}) + t_ap_a(1-p_a)}{(t_{a-1}+t_a)^2} \qquad \square$$

El teorema 4, empleando nuevamente el método delta, identifica la distribución de $\mathrm{logit}\left(\widehat{p^*}_{a-1}\right)$ requerida.

Teorema 4. *Si $\widehat{p^*}_{a-1}$ se distribuye como en el teorema 3, entonces:*

$$logit\left(\widehat{p^*}_{a-1}\right) \sim AN\left(\mu^*_{a-1}, (\sigma^2)^*_{a-1}\right)$$

donde:

$$\begin{aligned}\mu^*_{a-1} &= logit\left(\frac{t_{a-1}p_{a-1} + t_ap_a}{t_{a-1}+t_a}\right), \quad y \\ (\sigma^2)^*_{a-1} &= \frac{t_{a-1}p_{a-1}(1-p_{a-1}) + t_ap_a(1-p_a)}{\left[t_{a-1}p_{a-1}(1-p_{a-1}) + t_ap_a(1-p_a) + \frac{t_{a-1}t_a}{t_{a-1}+t_a}(p_{a-1}-p_a)^2\right]^2}\end{aligned}$$

Demostración. Claramente la función logit es no nula y diferenciable, luego el método delta garantiza la normalidad asintótica con la esperanza señalada. Resta probar la expresión de la varianza asintótica, como sigue:

$$\begin{aligned}
\left(\sigma^2\right)^*_{a-1} &= \mathrm{V}[\widehat{p^*}_{a-1}] \left[\frac{\mathrm{d}\ \mathrm{logit}(p^*_{a-1})}{\mathrm{d}\ p^*_{a-1}}\right]^2 \\
&= \frac{[t_{a-1}p_{a-1}(1-p_{a-1}) + t_a p_a(1-p_a)]}{(t_{a-1}+t_a)^2} \left[\frac{1}{p^*_{a-1}(1-p^*_{a-1})}\right]^2 \\
&= \frac{\mathrm{V}[Y^*_{a-1}]}{\left(\mathrm{V}_{\mathrm{Bin}}[Y^*_{a-1}]\right)^2} = \frac{\mathrm{V}[Y^*_{a-1}]}{\left(\mathrm{V}[Y^*_{a-1}] + \Delta V\right)^2} \\
&= \frac{t_{a-1}p_{a-1}(1-p_{a-1}) + t_a p_a(1-p_a)}{\left[t_{a-1}p_{a-1}(1-p_{a-1}) + t_a p_a(1-p_a) + \frac{t_{a-1}t_a}{t_{a-1}+t_a}(p_{a-1}-p_a)^2\right]^2}
\end{aligned}$$

□

Para efectos de las pruebas de hipótesis e intervalos de confianza deseados, los parámetros distribucionales se sustituyen por sus estimadores. Así, el parámetro poblacional p_i se sustituye por su estimador $\widehat{p}_i$ y el vector de parámetros $\boldsymbol{\beta}$ en la predictora lineal, por su estimador $\widehat{\boldsymbol{\beta}}$. Ambos calculados a partir de la muestra observada por el método de Newton-Raphson, como se señaló al inicio de esta sección. Consecuentemente, el procedimiento propuesto en la situación de agregación de niveles del factor es el siguiente:

1. Ajustar un modelo logit (llámese M) sobre los datos como se disponen en la tabla 1. Preservar del cómputo el vector de estimaciones de los p_i y la matriz estimada W.

2. Calcular la estimación puntual de p^*_{a-1} como en (7).

3. Calcular $\mathrm{logit}(\widehat{p}_i), i = 1, \ldots, a-2$ y $\mathrm{logit}\left(\widehat{p^*}_{a-1}\right)$ formando el vector

$$\mathrm{logit}(\widehat{p}^*)_{(a-1)\times 1}$$

4. Estimar la nueva matriz asintótica de varianzas y covarianzas para la situación del modelo saturado como:

$$\Sigma = \begin{bmatrix}
\left[t_1\widehat{p}_1(1-\widehat{p}_1)\right]^{-1} & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \left[t_{a-2}\widehat{p}_{a-2}(1-\widehat{p}_{a-2})\right]^{-1} & 0 \\
0 & \cdots & 0 & \widehat{(\sigma^2)}^*_{a-1}
\end{bmatrix}$$

donde $\widehat{(\sigma^2)}^*_{a-1}$ es la varianza obtenida en el teorema 4, estimada a partir de la muestra, sustituyendo p_i por $\widehat{p}_i$. La matriz Σ se conforma con la matriz W^{-1} del ajuste logit original. Se trata de la matriz de varianzas y covarianzas del vector columna cuyas componentes son los $\mathrm{logit}(\widehat{p}_i)$, $i = 1, \ldots, a-2$ [como se puede apreciar en la ecuación (5)], sustituyendo sus dos últimos elementos en

la diagonal por la varianza adecuada cuando se agregan los correspondientes dos últimos niveles del factor. Σ resulta diagonal pues las respuestas para los nuevos $a-1$ niveles no han perdido su condición de independencia asintótica.

5. Construir la nueva matriz de diseño $X^*_{(a-1)\times r^*}$ según el nuevo vector de parámetros deseado $\boldsymbol{\beta}^*_{r^*\times 1}$. Para el modelo saturado en (2), $r^* = a-1$.

6. Ajustar una regresión de mínimos cuadrados generalizados (Christensen 2002, pp. 34 y 88) con un nuevo modelo (llámese M^*) formulado como sigue:

$$Y = \text{logit}(\widehat{p^*}) = X^*\boldsymbol{\beta^*} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{AN}(0, \Sigma) \tag{8}$$

Ahora bien, en la situación estudiada, considerando la parametrización de referencia, el modelo saturado y en consecuencia la existencia de la matriz $(X^*)^{-1}$, los cómputos se simplifican notablemente como sigue:

$$\widehat{\beta}^* = (X^*)^{-1}Y$$
$$\text{V}\left[\widehat{\boldsymbol{\beta}}^*\right] = (X^*)^{-1}\Sigma\left[(X^*)^{-1}\right]'$$

Adicionalmente, el procedimiento sugerido presenta una ventaja computacional frente al procedimiento habitual: como se ha mencionado, al ajustar un modelo logit en general se recurre al método de Newton-Raphson. Este método implica una serie de iteraciones hasta alcanzar la convergencia en la estimación de los parámetros. En cada iteración es necesario invertir la matriz $X'WX$ (si el modelo es no saturado) o bien la matriz W (si el modelo es saturado). Luego, al ejecutar el procedimiento de ajuste del modelo logit en dos ocasiones, se duplica el esfuerzo computacional dedicado a la inversión de matrices. Por otra parte, siguiendo el procedimiento sugerido, es necesario dedicar tiempo de cómputo a la inversión de matrices una sola vez, pues el segundo ajuste surge aprovechando los resultados obtenidos en la primera oportunidad, sin necesidad de iterar nuevamente.

5. Extensión al caso cuando se agrupa un número cualquiera k de niveles $(1 < k < a)$

Para efectos de simplificar la exposición, el procedimiento sugerido considera la agrupación de los dos últimos niveles del factor explicativo. No obstante, es sencillo observar que dicho procedimiento puede extenderse al caso cuando el investigador decide agrupar k $(1 < k < a)$ niveles. Evidentemente, sigue en pie la violación del supuesto distribucional, cuando al menos dos de las probabilidades de éxito involucradas en la agrupación son distintas entre sí. Por otra parte, nótese que la posición que ocupan en la tabla los distintos niveles del factor es irrelevante (pues para el caso pueden ser reacomodados). Entonces, sin pérdida de generalidad, sean los últimos niveles $a-k+1, a-k+2, \ldots, a$ aquellos que el investigador decide agrupar, formando la nueva variable aleatoria $Y^*_{a-k+1} = Y_{a-k+1} + Y_{a-k+2} + \cdots + Y_a$.

Para simplificar los recorridos, sea $\nu = a - k + 1$. El teorema 5 establece la diferencia entre las varianzas binomial ($\mathrm{V}_{\mathrm{Bin}}[Y_\nu^*]$) y correcta ($\mathrm{V}[Y_\nu^*]$) para este caso más general.

Teorema 5. *Sean* $V_{Bin}[Y_\nu^*] = t_\nu^* p_\nu^* (1 - p_\nu^*)$ *y* $V[Y_\nu^*] = \sum_{i=\nu}^{a} t_i p_i (1 - p_i)$*, donde:*

$$t_\nu^* = \sum_{i=\nu}^{a} t_i \qquad y \qquad p_\nu^* = \frac{E[Y_\nu^*]}{t_\nu^*} = \frac{\sum_{i=\nu}^{a} t_i p_i}{\sum_{i=\nu}^{a} t_i}$$

Entonces:

$$\Delta V_\nu = V_{Bin}[Y_\nu^*] - V[Y_\nu^*] = \frac{\sum_{i=\nu}^{a-1} \sum_{j=i+1}^{a} t_i t_j (p_i - p_j)^2}{\sum_{i=\nu}^{a} t_i} \tag{9}$$

Demostración.

$$\Delta V_\nu = \left(\sum_{i=\nu}^{a} t_i \right) \left(\frac{\sum_{i=\nu}^{a} t_i p_i}{\sum_{i=\nu}^{a} t_i} \right) \left(1 - \frac{\sum_{i=\nu}^{a} t_i p_i}{\sum_{i=\nu}^{a} t_i} \right) - \sum_{i=\nu}^{a} t_i p_i (1 - p_i)$$

luego

$$\begin{aligned}
\left(\sum_{i=\nu}^{a} t_i \right) \Delta V_\nu &= \left(\sum_{i=\nu}^{a} t_i p_i \right) \left(\sum_{i=\nu}^{a} t_i - \sum_{i=\nu}^{a} t_i p_i \right) - \left(\sum_{i=\nu}^{a} t_i \right) \left(\sum_{i=\nu}^{a} t_i p_i (1 - p_i) \right) \\
&= \left(\sum_{i=\nu}^{a} t_i \right) \left(\sum_{i=\nu}^{a} t_i p_i^2 \right) - \left(\sum_{i=\nu}^{a} t_i p_i \right)^2 \\
&= \sum_{i=\nu}^{a} (t_i p_i)^2 + \sum_{i=\nu}^{a} t_i \sum_{\substack{j=\nu \\ j \neq i}}^{a} t_j p_j^2 - \left(\sum_{i=\nu}^{a} (t_i p_i)^2 + \sum_{i=\nu}^{a} t_i p_i \sum_{\substack{j=\nu \\ j \neq i}}^{a} t_j p_j \right) \\
&= \sum_{i=\nu}^{a} t_i \sum_{\substack{j=\nu \\ j \neq i}}^{a} t_j p_j^2 - \sum_{i=\nu}^{a} t_i p_i \sum_{\substack{j=\nu \\ j \neq i}}^{a} t_j p_j = \sum_{i=\nu}^{a} \sum_{\substack{j=\nu \\ j \neq i}}^{a} t_i t_j p_j (p_j - p_i)
\end{aligned}$$

$$
\begin{aligned}
&= t_\nu t_{\nu+1} p_{\nu+1}(p_{\nu+1} - p_\nu) + t_\nu t_{\nu+2} p_{\nu+2}(p_{\nu+2} - p_\nu) + \cdots + t_\nu t_a p_a (p_a - p_\nu) \\
&+ t_{\nu+1} t_\nu p_\nu (p_\nu - p_{\nu+1}) + t_{\nu+1} t_{\nu+2} p_{\nu+2}(p_{\nu+2} - p_{\nu+1}) + \cdots \\
&\qquad\qquad + t_{\nu+1} t_a p_a (p_a - p_{\nu+1}) \\
&+ t_{\nu+2} t_\nu p_\nu (p_\nu - p_{\nu+2}) + t_{\nu+2} t_{\nu+1} p_{\nu+1}(p_{\nu+1} - p_{\nu+2}) + \cdots \\
&\qquad\qquad + t_{\nu+2} t_a p_a (p_a - p_{\nu+2}) \\
&\vdots \\
&+ t_a t_\nu p_\nu (p_\nu - p_a) + t_a t_{\nu+1} p_{\nu+1}(p_{\nu+1} - p_a) + \cdots + t_a t_{a-1} p_{a-1}(p_{a-1} - p_a) \\
&= t_\nu t_{\nu+1}[p_{\nu+1}(p_{\nu+1} - p_\nu) + p_\nu(p_\nu - p_{\nu+1})] + t_\nu t_{\nu+2}[p_{\nu+2}(p_{\nu+2} - p_\nu) \\
&+ p_\nu(p_\nu - p_{\nu+2})] + \cdots + t_{a-1} t_a [p_a(p_a - p_{a-1}) + p_{a-1}(p_{a-1} - p_a)] \\
&= \sum_{i=\nu}^{a-1} \sum_{j=i+1}^{a} t_i t_j (p_i - p_j)^2 \\
\therefore \Delta V_\nu &= \frac{\displaystyle\sum_{i=\nu}^{a-1} \sum_{j=i+1}^{a} t_i t_j (p_i - p_j)^2}{\displaystyle\sum_{i=\nu}^{a} t_i}
\end{aligned}
$$

□

Así, en general también se sostiene que $\mathrm{V}_{\mathrm{Bin}}[Y_\nu^*] \geq \mathrm{V}[Y_\nu^*]$, ya que (9) es una cantidad no negativa. La extensión de los teoremas 3 y 4 es inmediata:

$$
\widehat{p^*}_\nu = \frac{\displaystyle\sum_{i=\nu}^{a} t_i \widehat{p}_i}{\displaystyle\sum_{i=\nu}^{a} t_i} \sim \mathrm{AN}\left(\frac{\displaystyle\sum_{i=\nu}^{a} t_i p_i}{\displaystyle\sum_{i=\nu}^{a} t_i}; \frac{\mathrm{V}[Y_\nu^*]}{\left(\displaystyle\sum_{i=\nu}^{a} t_i\right)^2} \right) \quad \text{y} \quad \mathrm{logit}\left(\widehat{p^*}_\nu\right) \sim \mathrm{AN}\left(\mu_\nu^*, (\sigma^2)_\nu^*\right)
$$

con:

$$
\mu_\nu^* = \mathrm{logit}\left(\frac{\displaystyle\sum_{i=\nu}^{a} t_i p_i}{\displaystyle\sum_{i=\nu}^{a} t_i} \right) \quad \text{y} \quad (\sigma^2)_\nu^* = \frac{\mathrm{V}[Y_\nu^*]}{(\mathrm{V}[Y_\nu^*] + \Delta V_\nu)^2}
$$

Consecuentemente, el procedimiento sugerido se extiende de la forma siguiente:

1. Ajustar un modelo logit sobre los datos originales, preservando el vector de estimaciones de los p_i y la matriz W.

2. Calcular $\widehat{p^*}_\nu$.

3. Calcular $\mathrm{logit}\left(\widehat{p}_i\right), i = 1, \ldots, a-k$ y $\mathrm{logit}\left(\widehat{p^*}_\nu\right)$ formando el vector $\mathrm{logit}\left(\widehat{p}^*\right)_{(\nu \times 1)}$.

4. Estimar la nueva matriz asintótica de covarianzas como:

$$\Sigma_\nu = \begin{bmatrix} \left[t_1\widehat{p}_1(1-\widehat{p}_1)\right]^{-1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \left[t_{a-k}\widehat{p}_{a-k}(1-\widehat{p}_{a-k})\right]^{-1} & 0 \\ 0 & \cdots & 0 & \widehat{(\sigma^2)}_\nu^* \end{bmatrix}$$

5. Construir la nueva matriz de diseño $X^*_{\nu\times r^*}$ según el nuevo vector de parámetros deseado $\beta^*_{r^*\times 1}$. Para el modelo saturado se tendrá que $r^* = a-k+1$.

6. Ajustar una regresión de mínimos cuadrados generalizados con un nuevo modelo formulado como: $Y = \text{logit}\big(\widehat{p^*}\big) = X^*\beta^* + \epsilon$, con $\epsilon \sim \text{AN}(0, \Sigma_\nu)$.

Y como antes, empleando la parametrización de referencia y el modelo saturado, $\widehat{\beta}^* = (X^*)^{-1}Y$ y $\text{V}\left[\widehat{\beta}^*\right] = (X^*)^{-1}\Sigma_\nu\left[(X^*)^{-1}\right]'$.

6. Ejemplo

En la tabla 2 se presenta una situación en que el interés se centra en estudiar la relación entre una variable respuesta Y y un factor explicativo A con tres niveles. Se muestran allí las frecuencias observadas para cada nivel.

Tabla 2: $Y(0,1)$ *vs.* $A(1,2,3)$.

	Y		
A	0	1	Total
1	189	161	350
2	300	50	350
3	32	318	350
Total	521	529	1050

El modelo logit saturado empleando la parametrización con β_3 como nivel de referencia, se muestra en la ecuación (10).

$$\begin{bmatrix} \text{logit}(p_1) \\ \text{logit}(p_2) \\ \text{logit}(p_3) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \tag{10}$$

La tabla 3 contiene las estimaciones de los parámetros de la predictora lineal con las pruebas χ^2 de Wald para $\text{H}_0 : \beta_i = 0, i = 1,2,3$ e intervalos de confianza (IC) del 95 % construidos para β_i.

Las probabilidades predichas y sus intervalos de confianza siguiendo a Agresti (2007, p. 109) se muestran en la tabla 4.

Ahora bien, supóngase que se agrupan los niveles 2 y 3 del factor A en la tabla 2 y se repite el procedimiento para el modelo saturado. En este caso, el nuevo

Tabla 3: Modelo original. $\widehat{\beta}_i$ y prueba de Wald ($H_0 : \beta_i = 0$).

	Estimación de β_i					IC del 95 %	
i	$\widehat{\beta}_i$	SE	χ^2	p-Valor	Conclusión	Li	Ls
1	2.296	0.185	153.3	< 0.0001	Rechazar	1.933	2.660
2	-2.457	0.214	131.5	< 0.0001	Rechazar	-2.877	-2.037
3	-4.088	0.240	289.5	< 0.0001	Rechazar	-4.559	-3.617

SE: Error Estándar Li: Límite inferior Ls: Límite superior

Tabla 4: Modelo original. Probabilidades predichas e IC del 95 %.

i	$\widehat{p}_i$	Li	Ls
1	0.4600	0.4084	0.5125
2	0.1429	0.1100	0.1836
3	0.9086	0.8736	0.9346

modelo es:

$$\begin{bmatrix} \text{logit}(p_1^*) \\ \text{logit}(p_2^*) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} \tag{11}$$

Las nuevas estimaciones utilizando el proceso habitual de ajuste, esto es, volviendo a ajustar un modelo logit sobre la tabla de contingencia resultante, se muestran en la tabla 5. Las probabilidades predichas por el modelo, sin atención a los niveles de significación de los parámetros, se reproducen en la tabla 6.

Tabla 5: Procedimiento habitual. $\widehat{\beta^*}_i$ y prueba de Wald ($H_0 : \beta_i^* = 0$).

	Estimación de β_i^*					IC del 95 %	
i	$\widehat{\beta^*}_i$	SE	χ^2	p-Valor	Conclusión	Li	Ls
1	0.103	0.076	1.850	0.174	No rechazar	-0.045	0.251
2	-0.263	0.131	4.023	0.045	Rechazar	-0.521	-0.006

Tabla 6: Procedimiento habitual. Probabilidades predichas e IC del 95 % sin considerar la significación de los parámetros del modelo.

i	$\widehat{p^*}_i$	Li	Ls
1	0.4600	0.4084	0.5125
2	0.5257	0.4887	0.5625

Como es de esperarse, en esta oportunidad el nuevo vector de parámetros β^* resulta estimado de forma diferente a la anterior. Por su parte, el nuevo vector de probabilidades predichas $\left(\widehat{p^*}\right)$, sin poner atención a los niveles de significación de los parámetros, resulta igual al anterior en la primera componente $\left(\widehat{p^*}_1 = \widehat{p}_1 = 0.460\right)$, y diferente, pero tal como se supone debería ocurrir, en la segunda componente. Esto es:

$$\widehat{p^*}_2 = \frac{t_2\widehat{p}_2 + t_3\widehat{p}_3}{t_2 + t_3} = \frac{350 \times 0.1429 + 350 \times 0.9086}{2 \times 350} = 0.526$$

Sin embargo, con $\alpha = 0.05$, en la tabla 5 se sugiere la inexistencia de evidencia suficiente para rechazar la hipótesis de nulidad de β_1^*, ya que el IC contiene al cero.

Esto tiene implicaciones importantes para el análisis: al no poderse concluir que β_1^* es significativamente diferente de 0, el modelo (11) pierde vigencia y las probabilidades predichas en la tabla 6, en estricto sentido estadístico, no deben considerarse válidas. Las predicciones correctas son entonces notablemente diferentes:

$$\widehat{p^*}_1 = \frac{e^{\widehat{\beta^*}_1+\widehat{\beta^*}_2}}{1+e^{\widehat{\beta^*}_1+\widehat{\beta^*}_2}} = \frac{e^{\widehat{\beta^*}_2}}{1+e^{\widehat{\beta^*}_2}} = 0.4346 \tag{12}$$

$$\widehat{p^*}_2 = \frac{e^{\widehat{\beta^*}_1}}{1+e^{\widehat{\beta^*}_1}} = \frac{e^0}{1+e^0} = 0.5 \tag{13}$$

Finalmente, la tabla 7 contiene las estimaciones del modelo logit en (11) y los IC del 95 %, y la tabla 8 las probabilidades predichas, ahora ajustando los datos según el procedimiento sugerido en este trabajo.

Tabla 7: Procedimiento sugerido. $\widehat{\beta^*}_i$ y prueba de Wald ($H_0 : \beta_i^* = 0$).

	Estimación de β_i^*					IC del 95 %	
i	$\widehat{\beta^*}_i$	SE	χ^2	p-Valor	Conclusión	Li	Ls
1	0.103	0.0486	4.49	0.034	Rechazar	0.0077	0.1982
2	−0.263	0.1177	5.00	0.025	Rechazar	−0.4941	−0.0325

Tabla 8: Procedimiento sugerido. Probabilidades predichas e IC del 95 %.

i	$\widehat{p^*}_i$	Li	Ls
1	0.4600	0.4084	0.5125
2	0.5257	0.5019	0.5494

Nótese que, tal como se esperaba, las estimaciones puntuales del procedimiento habitual y del procedimiento sugerido son en esencia las mismas. No obstante, las varianzas estimadas son diferentes en ambos procedimientos. Las varianzas estimadas por el procedimiento sugerido son menores (y por tanto preferibles) que las estimadas mediante el procedimiento habitual, tanto en lo que se refiere a los estimadores de los parámetros de la predictora lineal como en cuanto a la segunda componente del vector de probabilidades.

Adicionalmente, ahora se produce un cambio en la conclusión sobre la significación de β_1^*. Mientras que por el procedimiento habitual las estimaciones estadísticamente válidas, ecuaciones (12) y (13), lucen considerablemente por debajo de lo que se supondría, mediante el procedimiento sugerido sí resultan estadísticamente válidas las predicciones de la tabla 8, que se aproximan de mejor manera a lo que cabría esperar con los datos disponibles.

Por último, la tabla 9 sintetiza los resultados obtenidos para efectos de su comparación con base en la varianza estimada. La comparación de los IC en términos absolutos carece de sentido sin considerar las estimaciones puntuales correspondientes, de manera que, para poder comparar las diferencias en las longitudes de los IC, se utiliza la razón entre las longitudes del primero y el segundo (obtenidos por el procedimiento habitual y sugerido, respectivamente), esto es, el cociente

entre las longitudes de los intervalos del 95 % de confianza, obtenidos por ambos métodos.

Tabla 9: Habitual *vs.* sugerido. Razón de la longitud de los IC.

	$\widehat{\beta^*}_i$			$\widehat{p^*}_i$		
i	$L(IC_h)$	$L(IC_s)$	$R_\beta = \frac{L(IC_h)}{L(IC_s)}$	$L(IC_h)$	$L(IC_s)$	$R_p = \frac{L(IC_h)}{L(IC_s)}$
1	0.2960	0.1905	1.5538	0.1041	0.1041	1.0000
2	0.5150	0.4616	1.1157	0.0738	0.0475	1.5537

$L(\cdot)$: Longitud. R: Razón.

Subíndices, h: Procedimiento habitual, s: Procedimiento sugerido.

De la tabla 9 se deduce que para $\widehat{\beta^*}_1$, el IC calculado por el procedimiento habitual es 1.5538 veces el calculado por el método sugerido. Para $\widehat{\beta^*}_2$, el primero es 1.1157 veces el segundo. Esto significa una mejora sustancial en la precisión de las estimaciones empleando el procedimiento sugerido, como cabría esperar luego del ajuste hacia la baja en el supuesto sobre la varianza de las muestras involucradas. En efecto, en este ejemplo particular, sea s_h^2 la estimación de la varianza calculada mediante el procedimiento habitual y $\widehat{\Delta V}$ la estimación de su diferencia respecto a la varianza correcta (teorema 2). Entonces:

$$s_h^2 = t_2^* \widehat{p^*}_2 (1 - \widehat{p^*}_2) = 700 \times 0.5257 \times (1 - 0.5257) = 174,54$$

$$\widehat{\Delta V} = \frac{t_2 t_3 (\widehat{p}_2 - \widehat{p}_3)^2}{t_2 + t_3} = \frac{350 \times (0.1429 - 0.9086)^2}{2} = 102,60$$

Así, $\widehat{\Delta V}$ estima una disminución en la varianza al ajustar el modelo según el procedimiento sugerido del $100 \times (174.54 - 102.60)/174.54 = 41.22\,\%$, que se refleja en una disminución de $100 \times (0.2960 - 0.1905)/0.2960 = 35.64\,\%$ en la longitud del IC para $\widehat{\beta^*}_1$ y en una disminución de $100 \times (0.5150 - 0.4616)/0.5150 = 10.37\,\%$ en la longitud del IC para $\widehat{\beta^*}_2$.

Nótese, además, que tal mejora no afecta la predicción de $\widehat{p^*}_1$, pero sí lo hace en la de $\widehat{p^*}_2$. De hecho, el efecto relativo en la disminución de la longitud del IC para $\widehat{p^*}_2$ es esencialmente el mismo que opera sobre $\widehat{\beta^*}_1$, en total acuerdo con lo esperado, puesto que en el modelo se postula una relación exclusiva entre estos dos parámetros ($\text{logit}(p_2^*) = \beta_1^*$).

7. Comparación de ambos procedimientos mediante simulación

La simulación cuyos resultados se comentan a continuación, ha sido diseñada para estudiar el efecto de la agregación de los niveles $a - 1$ y a del factor explicativo, mediante la generación seudoaleatoria de un número grande de tablas de contingencia, como la tabla 1, y su posterior análisis sobre medidas-resumen de desempeño. Para hacer más simple el experimento, se sigue la estructura del ejemplo de la sección anterior, con solo tres niveles del factor, agrupando los dos últimos (2 y 3).

7.1. Diseño del experimento de simulación

A la luz del examen de las varianzas presentado en la sección 3, con el propósito de amplificar el impacto de la diferencia entre ambos procedimientos, el experimento supone $t_1 = t_2 = t_3 = t = 350$. Como en la situación no luce de interés la comparación del efecto de ambos procedimientos sobre el primer nivel del factor, para este se genera seudoaleatoriamente p_1 a partir de una distribución uniforme en $(0, 1)$. Con los valores de t y p_1 se genera la muestra $Y_1 \sim \text{Bin}(t, p_1)$. Para los restantes niveles del factor (objeto de comparación) se generan las muestras $Y_2 \sim \text{Bin}(t, p_2)$ y $Y_3 \sim \text{Bin}(t, p_3)$, para las combinaciones $\Delta p = |p_2 - p_3| = 0.0, 0.2, 0.4, 0.6, 0.8$, obtenidas manteniendo constante el valor de $p_2 = 0.1$ y variando el valor de $p_3 = 0.1, 0.3, 0.5, 0.7, 0.9$.

Para cada uno de los 5 valores a experimentar de Δp, se generan 2000 tablas de contingencia constituidas por binomiales independientes, dentro de cada tabla y entre tablas. El experimento prosigue ajustando un primer modelo logit contando los tres niveles del factor, un segundo modelo logit ajustado a la tabla resultante de agrupar los niveles 2 y 3 del factor, y un tercer ajuste de dicha tabla, mediante el procedimiento sugerido. El nivel de significación para las pruebas se establece en $\alpha = 0.05$.

Las medidas de desempeño que se emplean como resultado de la simulación se exponen a continuación:

a) En primer lugar, se utilizan las diferencias en las estimaciones puntuales de $\beta_1^*, \beta_2^*, p_1^*, p_2^*$, obtenidas mediante los procedimientos habitual y sugerido, sin tomar en cuenta la significación de los parámetros de la predictora lineal.

b) Para la comparación de las diferencias en las longitudes de los IC, calculados mediante los procedimientos habitual y sugerido, se utiliza el promedio de la razón entre las longitudes del primero y el segundo. Estas razones se calculan para los IC que acompañan a las estimaciones de los parámetros $\beta_1^*, \beta_2^*, p_1^*, p_2^*$.

c) Finalmente, se estudian las frecuencias absolutas de ocurrencia del cambio en la conclusión del análisis de varianza (aceptación a rechazo, o viceversa) para las pruebas de hipótesis sobre los parámetros $H_0 : \beta_1^* = 0$ y $H_0 : \beta_2^* = 0$, cuando se los contrasta por el procedimiento habitual, y cuando se los contrasta por el procedimiento sugerido.

7.2. Resultados del experimento de simulación

a) Cuando no se toma en cuenta la significación de los parámetros de la predictora lineal, en cuanto a las estimaciones puntuales, redondeando al tercer decimal, en todos los casos estas coinciden por ambos métodos.

   Así, en cada una de las 10000 tablas de contingencia generadas, los estimadores $\widehat{\beta^*}_1$, $\widehat{\beta^*}_2$, $\widehat{p^*}_1$ y $\widehat{p^*}_2$ se calculan esencialmente de igual forma cuando se les estima por el procedimiento habitual o por el procedimiento sugerido.

b) En cuanto a las longitudes de los IC para cada estimador, la tabla 10 presenta los resultados de las razones promedio obtenidas.

Tabla 10: Razón de la longitud de los IC para los estimadores.

	$\widehat{\beta^*}_1$		$\widehat{\beta^*}_2$		$\widehat{p^*}_1$		$\widehat{p^*}_2$	
Δp	Med.	D.E.	Med.	D. E.	Med.	D. E.	Med.	D. E.
0.0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.2	1.03	0.01	1.01	0.00	1.00	0.00	1.03	0.01
0.4	1.11	0.02	1.03	0.01	1.00	0.00	1.11	0.02
0.6	1.27	0.04	1.06	0.01	1.00	0.00	1.27	0.04
0.8	1.68	0.08	1.10	0.02	1.00	0.00	1.67	0.09

Med.: media. D. E.: desviación estándar.

Nótese en la tabla 10 que a mayor Δp, tanto mayor es la razón promedio de las longitudes de los IC estimados por ambos métodos. Por ejemplo, la longitud del IC correspondiente a $\widehat{\beta^*}_1$ estimado por el procedimiento habitual, es en promedio 1.68 veces la del IC estimado por el método sugerido, cuando $\Delta p = |p_2 - p_3| = 0.8$. Por otro lado, es claro que cuando las probabilidades de éxito de las muestras agrupadas están cercanas, ambos métodos estiman en promedio esencialmente los mismos IC para todos los parámetros. Este hecho debía verificarse en virtud de los postulados del teorema 2. También resulta que el comportamiento relativo de los IC es muy similar para $\widehat{\beta^*}_1$ y $\widehat{p^*}_2$, en total concordancia con lo esperado, tal y como se mencionó en la sección anterior. Por último, es apreciable que los IC correspondientes a $\widehat{p*}_1$, en promedio no muestran diferencias al ser estimados por ambos métodos, mientras que las razones promedio de las longitudes de los IC correspondientes a los $\widehat{\beta^*}_2$, aunque también muestran un comportamiento creciente con Δp, lo hacen en proporción considerablemente menor que para $\widehat{\beta^*}_1$.

c) Finalmente, la tabla 11 muestra las frecuencias absolutas de ocurrencia de las conclusiones sobre los parámetros del modelo. Para β_1^* y Δp por debajo de 0.8, ambos procedimientos conducen al rechazo de la hipótesis de nulidad en todas las muestras. Para $\Delta p = 0.8$, la situación es contraria, es decir, en la mayoría de las muestras no hay evidencia suficiente para rechazar la hipótesis de nulidad. No obstante, mientras que utilizando el procedimiento habitual se rechaza la hipótesis nula en 3 de las 2000 muestras $(0,15\,\%)$, empleando el procedimiento sugerido esto ocurre en 109 de las 2000 muestras $(5.46\,\%)$.

Tabla 11: Frecuencias en las conclusiones sobre los parámetros.

	$H_0 : \beta_1^* = 0$				$H_0 : \beta_2^* = 0$			
	Proc. hab.		Proc. sug.		Proc. hab.		Proc. sug.	
Δp	Acep.	Rech.	Acep.	Rech.	Acep.	Rech.	Acep.	Rech.
0.0	0	2000	0	2000	104	1896	104	1896
0.2	0	2000	0	2000	285	1715	283	1717
0.4	0	2000	0	2000	275	1725	271	1729
0.6	0	2000	0	2000	291	1709	274	1726
0.8	1997	3	1891	109	302	1698	268	1732

Al disminuir la región de aceptación de la hipótesis nula, mejorando la precisión en la estimación del parámetro y la potencia de la prueba de la hipótesis de nulidad, el procedimiento sugerido favorece la investigación particular que se esté realizando. La calidad de las conclusiones sobre las posibilidades y sobre las probabilidades mejora siempre que sea estadísticamente válido el rechazo de la hipótesis de nulidad de los parámetros.

Para β_2^* la situación es variada. Aunque se observa para todos los valores de Δp una alta proporción de rechazo a la hipótesis de nulidad del parámetro, leve pero sostenidamente a medida que aumenta Δp, el procedimiento sugerido favorece cada vez más el rechazo de tal hipótesis, cuando se lo compara con el procedimiento habitual. Sin embargo, para $\Delta p = 0.8$, esto es, la máxima diferencia entre las probabilidades de éxito estudiadas por el experimento, el procedimiento sugerido conduce a cambiar la conclusión en apenas $1732 - 1698 = 34$ oportunidades más que el procedimiento habitual, es decir, cerca de tres veces menos que lo ocurrido para β_1^* $(109 - 3 = 106)$.

En el anexo se incluye el código fuente R (R Development Core Team 2007) programado para realizar la simulación.

8. Conclusiones

Es clara la existencia de un problema en el ajuste del modelo logit con niveles agrupados del factor. Se sugiere en este trabajo un procedimiento que, aprovechando la disminución en la varianza cuando se postula el modelo distribucional correcto en lugar del modelo binomial, conduce a la reducción de los errores estándares de las estimaciones en un porcentaje apreciable.

Al disminuir el error estándar estimado, disminuye la región de aceptación de la hipótesis de nulidad de los parámetros de la predictora lineal, mejorando las estimaciones y la potencia de la prueba de dicha hipótesis de nulidad. Consecuentemente, el procedimiento sugerido resulta preferible al habitual.

Desde el punto de vista computacional, el procedimiento sugerido resulta más eficiente que el habitual puesto que, aprovechando los cómputos obtenidos en un primer ajuste del modelo logit, produce nuevas estimaciones mediante regresión, sin requerir más iteraciones.

Mediante simulación se ha corroborado que diferencias relativamente leves en las probabilidades de éxito de las muestras involucradas en la agregación, en promedio no conducen a concluir que la aplicación del procedimiento sugerido sea diferente frente a la aplicación del habitual. Esto muestra clara evidencia sobre la robustez del modelo logit, sus habituales procedimientos de ajuste y prueba de hipótesis. Por otra parte, a medida que dicha diferencia en las probabilidades de éxito se profundiza (para los valores estudiados, esto es $\Delta p > 0.6$), resulta cada vez mejor apoyarse en el procedimiento sugerido. Consecuentemente, el investigador debe tener especial cuidado cuando agrupa niveles del factor, cuyas proporciones muestrales son notoriamente disímiles.

En este trabajo se ilustró la situación considerando un solo factor, la parametrización de referencia y el modelo saturado. La investigación continuará en varias direcciones: asumiendo la presencia de dos o más factores, postulando el modelo no saturado para su análisis, estudiando el efecto sobre los residuos y las medidas de bondad del ajuste, entre otras.

Agradecimientos

Los autores agradecen al Consejo de Desarrollo Científico, Humanístico y Tecnológico (CDCHT) de la Universidad de Los Andes el apoyo financiero brindado para la realización de este trabajo. Asimismo, agradecen los valiosos comentarios de los árbitros anónimos, los cuales sin duda contribuyeron a mejorar tanto el fondo como la forma del presente trabajo.

[Recibido: junio de 2009 — Aceptado: agosto de 2009]

Referencias

Agresti, A. (2007), *An Introduction to Categorical Data Analysis*, 2 edn, John Wiley & Sons, Inc., New Jersey, United States.

Christensen, R. (1997), *Log-Linear Models and Logistic Regression*, 2 edn, Springer-Verlag, New York, United States.

Christensen, R. (2002), *Plane Answers to Complex Questions. The Theory of Linear Models*, 3 edn, Springer-Verlag, New York, United States.

Collett, D. (2002), *Modelling binary data*, 2 edn, Chapman & Hall/CRC, Boca Raton, United States.

Cox, D. R. (1970), *Analysis of Binary Data*, 1 edn, Methuen and Co Ltd., London, England.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3 edn, John Wiley & Sons. Inc., New York, United States.

Grizzle, J. E., Starmer, C. F. & Koch, G. G. (1969), 'Analysis of Categorical Data by Linear Models', *Biometrics* **25**(3), 489–504.

Hilbe, J. M. (2009), *Logistic Regression Models*, 1 edn, Chapman & Hall, Florida, United States.

Hodges, J. L. & Le Cam, L. (1960), 'The Poisson Approximation to the Poisson Binomial Distribution', *The Annals of Mathematical Statistics* **31**(3), 737–740.

Hosmer, D. W. & Lemeshow, S. (2000), *Applied Logistic Regression*, 2 edn, John Wiley & Sons, New York, United States.

Lehmann, E. L. (1999), *Elements of Large-Sample Theory*, 1 edn, Springer-Verlag, New York, United States.

McCullagh, P. & Nelder, J. (1989), *Generalized Linear Models*, 2 edn, Chapman & Hall, London, England.

McCulloch, C. E. & Searle, S. R. (2001), *Generalized, Linear, and Mixed Models*, 1 edn, John Wiley & Sons, Inc., New York, United States.

Neammanee, K. (2005), 'A refinement of Normal approximation to Poisson Binomial', *International Journal of Mathematics and Mathematical Sciences* (5), 717–728.

Nedelman, J. & Wallenius, T. (1986), 'Bernoulli Trials, Poisson Trials, Surprising Variances, and Jensen's Inequality', *The American Statistician* **40**(4), 286–289.

Nelder, J. A. & Wedderburn, R. W. M. (1972), 'Generalized Linear Models', *Journal of the Royal Statistical Society. Serie A* (135), 370–384.

Neyman, J. (1939), 'On a new class of contagious distributions, applicable in entomology and bacteriology', *The Annals of Mathematical Statistics* **10**(1), 35–57.

Ollero, H. J. & Ramos, R. H. M. (1991), 'La distribución hipergeométrica como binomial de poisson', *Trabajos de Estadística* **6**(1), 35–43.

Ponsot, E. (2009), Estudio de la agrupación de niveles en el modelo logit, tesis de doctorado, Instituto de Estadística Aplicada y Computación, Facultad de Ciencias Económicas y Sociales, Universidad de Los Andes, Mérida, Venezuela.

Powers, D. A. & Xie, Y. (1999), *Statistical Methods for Categorical Data Analysis*, 1 edn, Academic Press, United States.

R Development Core Team (2007), 'R: A language and environment for statistical computing', Vienna, Austria.
*http://www.R-project.org

Rodríguez, G. (2008), 'Lectures notes about generalized linear models', New Jersey, United States.
*http://data.princeton.edu/wws509/notes

Rohatgi, V. & Ehsanes, A. (2001), *An Introduction to Probability and Statistics*, 2 edn, John Wiley & Sons, Inc., New York, United States.

Roos, B. (1999), 'Asymptotics and Sharp Bounds in the Poisson Approximation to the Poisson-Binomial Distribution', *Bernoulli* **5**(6), 1021–1034.

Sprott, D. A. (1958), 'The Method of Maximum Likelihood Applied to the Poisson Binomial Distribution', *Biometrics* **14**(1), 97–106.

Wang, Y. H. (1993), 'On the Number of Successes in Independent Trials', *Statistica Sinica* **3**, 295–312.

Weba, M. (1999), 'Bounds for the Total Variation Distance between the Binomial and the Poisson Distribution in case of Medium-Sized Success Probabilities', *Journal of Applied Probability* (36), 497–104.

Wedderburn, R. W. M. (1974), 'Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method', *Biometrika* **61**(3), 439–447.

Apéndice A. Código R para la simulación

```
require(binom)
require(vcd)
require(stats)
logit <- function(p) {
   Resulta<-log(p/(1-p))
}
set.seed(1055047082)
t=350
N=10000
Datos <- array(0, dim=c(N,12))
p2=0.1
p3 <- c(0.1,0.3,0.5,0.7,0.9)
j <- 1
for (i in 1:N) {
   Datos[i,1] <-i;
   Datos[i,2] <-round(runif(1,0.1,0.9),1);
   Datos[i,3] <-p2;
   Datos[i,4] <-p3[j];
   Datos[i,5] <-abs(Datos[i,3]-Datos[i,4])
   Datos[i,6] <-t;
   Datos[i,7] <-rbinom(1,Datos[i,6],Datos[i,2]);
   Datos[i,8] <-rbinom(1,Datos[i,6],Datos[i,3]);
   Datos[i,9] <-rbinom(1,Datos[i,6],Datos[i,4]);
   Datos[i,10]<-Datos[i,6]-Datos[i,7];
   Datos[i,11]<-Datos[i,6]-Datos[i,8];
   Datos[i,12]<-Datos[i,6]-Datos[i,9];
   j<-j+1;
   if (j>5) j=1
}
TabDatos <- data.frame(Id=Datos[,1], p1=Datos[,2], p2=Datos[,3], p3=Datos[,4],
      Dp=Datos[,5], t=Datos[,6], y11=Datos[,7], y12=Datos[,8],
      y13=Datos[,9], y01=Datos[,10], y02=Datos[,11], y03=Datos[,12])
write.table(TabDatos, file = "C:/Datos.txt",
      append = FALSE, dec = ".", quote = FALSE, sep = ";",
      row.names = FALSE, col.names = TRUE)
##### Nivel de significación (Alfa)
k<-qnorm(0.05/2, mean=0, sd=1, lower.tail = FALSE, log.p = FALSE)
##### Fin de nivel de significación
##### Inicia la simulación
Resulta <- array(0, dim=c(N,42))
for (i in 1:N) {
```

```
##### Datos
### Tabla original
Ao <-factor(c(1,2,3))
contrasts(Ao) <- contr.treatment(3,3)
Yo <- cbind(Y1=c(TabDatos$y11[i],TabDatos$y12[i],TabDatos$y13[i]),
      Y0=c(TabDatos$y01[i],TabDatos$y02[i],TabDatos$y03[i]))
to <- c(t,t,t)
dat_o <- data.frame(Ao, Yo)
### Tabla agrupando los niveles 2 y 3
A <- factor(Ao[1:2])
contrasts(A) <- contr.treatment(2,2)
Y <- Yo[1:2,1:2]
Y[2,1] <- Yo[2,1]+Yo[3,1]
Y[2,2] <- Yo[2,2]+Yo[3,2]
ta <- to[1:2]
ta[2]<-to[2]+to[3]
dat_a <- data.frame(A, Y)
##### Fin de datos
##### Modelo logit original
o_mod <- glm(Yo ~ Ao, family = binomial, data = dat_o, x = TRUE)
### Estimación de parámetros del modelo (Beta) e intervalos de confianza
o_b <- data.frame(cbind(b=summary(o_mod)$coefficients[,1],
   li=summary(o_mod)$coefficients[,1]-k*summary(o_mod)$coefficients[,2],
   ls=summary(o_mod)$coefficients[,1]+k*summary(o_mod)$coefficients[,2]))
### Predicción de probabilidades (p) según el modelo
o_p0 <- predict(o_mod, type = "response")
### Intervalos de confianza para p siguiendo a Agresti (2007:109)
o_Vb <- summary(o_mod)$cov.unscaled
o_p1 <- k*sqrt(diag(o_mod$x %*% o_Vb %*% t(o_mod$x)))
o_p2 <- data.frame(cbind(li=logit(o_p0)-o_p1, ls=logit(o_p0)+o_p1))
o_p  <- data.frame(cbind(p=o_p0, li=exp(o_p2$li)/(1+exp(o_p2$li)),
      ls=exp(o_p2$ls)/(1+exp(o_p2$ls))))
##### Fin del modelo logit original
##### Modelo logit agrupando niveles 2 y 3 (Procedimiento habitual)
h_mod <- glm(Y ~ A, family = binomial, data = dat_a, x = TRUE)
### Estimación de parámetros del modelo (Beta*) e intervalos de confianza
h_b <- data.frame(cbind(b=summary(h_mod)$coefficients[,1],
   li=summary(h_mod)$coefficients[,1]-k*summary(h_mod)$coefficients[,2],
   ls=summary(h_mod)$coefficients[,1]+k*summary(h_mod)$coefficients[,2]))
### Predicción de probabilidades (p*) según el modelo
h_p0 <- predict(h_mod, type = "response")
### Intervalos de confianza para p* siguiendo a Agresti (2007:109)
h_Vb <- summary(h_mod)$cov.unscaled
h_p1 <- k*sqrt(diag(h_mod$x %*% h_Vb %*% t(h_mod$x)))
h_p2 <- data.frame(cbind(li=logit(h_p0)-h_p1, ls=logit(h_p0)+h_p1))
h_p  <- data.frame(cbind(p=h_p0, li=exp(h_p2$li)/(1+exp(h_p2$li)),
      ls=exp(h_p2$ls)/(1+exp(h_p2$ls))))
##### Fin de modelo logit agrupando niveles 2 y 3 (procedimiento habitual)
#####################################################
##### Ajuste sugerido agrupando niveles 2 y 3 #####
#####################################################
### Estimación de parámetros del modelo (Beta*) e intervalos de confianza
s_x     <- o_mod$x[1:2,1:2]
s_x_svd <- svd(s_x)
s_xInv  <- s_x_svd$v %*% diag(1/s_x_svd$d) %*% t(s_x_svd$u)
### Predicción de logit(p*) a partir del modelo original
```

```
   s_p    <- as.vector(o_p$p[1:2])
   s_p[2] <- (to[2]*o_p$p[2]+to[3]*o_p$p[3])/(to[2]+to[3])
   s_lp   <- logit(s_p)
   ### Estimación de los beta
   s_b <- as.vector(s_xInv %*% s_lp)
   #print(s_p)
   ### Estimación de Sigma
   o_Vxb <- o_mod$x %*% o_Vb %*% t(o_mod$x)
   sigma <- o_Vxb[1:2,1:2]
   sigma[2,2] <-
      ((to[2]*o_p$p[2]*(1-o_p$p[2])+to[3]*o_p$p[3]*(1-o_p$p[3]))*(to[2]+to[3])^2) /
      ((to[2]*o_p$p[2]+to[3]*o_p$p[3])*(to[2]+to[3]-to[2]*o_p$p[2]-to[3]*o_p$p[3]))^2
   ### Intervalos de confianza para los Beta*
   s_Vb  <- s_xInv %*% sigma %*% t(s_xInv)
   s_SEb <- sqrt(diag(s_Vb))
   s_Chi <- s_b^2/s_SEb^2
   s_b   <- data.frame(cbind(b=s_b, SE=s_SEb, Chi2=s_Chi,
         Val_p=pchisq(s_Chi, df=1, lower.tail = FALSE),
         li=s_b-k*s_SEb, ls=s_b+k*s_SEb))
   ### Intervalos de confianza para p* siguiendo a Agresti (2007:109)
   s_SElp <- as.vector(sqrt(diag(sigma)))
   s_lp   <- data.frame(cbind(logit_p=s_lp, li=s_lp-k*s_SElp, ls=s_lp+k*s_SElp))
   s_p    <- data.frame(cbind(p=s_p, li=exp(s_lp$li)/(1+exp(s_lp$li)),
          ls=exp(s_lp$ls)/(1+exp(s_lp$ls))))
   ##### Fin de ajuste sugerido agrupando niveles 2 y 3
   ##### Guardando los resultados pertinentes (0=Se acepta Ho, 1=Se rechaza Ho)
   Resulta[i,1]=i
   Resulta[i,2]=Datos[i,5]
   Resulta[i,3] =h_b$b[1]; Resulta[i,4] =h_b$li[1]; Resulta[i,5] =h_b$ls[1]
   Resulta[i,6] =h_b$ls[1]-h_b$li[1]
   Resulta[i,7] =ifelse(h_b$ls[1]>=0 & h_b$li[1]<=0, 0, 1)
   Resulta[i,8] =s_b$b[1]; Resulta[i,9] =s_b$li[1]; Resulta[i,10]=s_b$ls[1]
   Resulta[i,11]=s_b$ls[1]-s_b$li[1]
   Resulta[i,12]=ifelse(s_b$ls[1]>=0 & s_b$li[1]<=0, 0, 1)
   Resulta[i,13]=Resulta[i,6]/Resulta[i,11]
   Resulta[i,14]=h_b$b[2]; Resulta[i,15]=h_b$li[2]; Resulta[i,16]=h_b$ls[2]
   Resulta[i,17]=h_b$ls[2]-h_b$li[2]
   Resulta[i,18]=ifelse(h_b$ls[2]>=0 & h_b$li[2]<=0, 0, 1)
   Resulta[i,19]=s_b$b[2]; Resulta[i,20]=s_b$li[2]; Resulta[i,21]=s_b$ls[2]
   Resulta[i,22]=s_b$ls[2]-s_b$li[2]
   Resulta[i,23]=ifelse(s_b$ls[2]>=0 & s_b$li[2]<=0, 0, 1)
   Resulta[i,24]=Resulta[i,17]/Resulta[i,22]
   Resulta[i,25]=h_p$p[1]; Resulta[i,26]=h_p$li[1]; Resulta[i,27]=h_p$ls[1]
   Resulta[i,28]=h_p$ls[1]-h_p$li[1]
   Resulta[i,29]=s_p$p[1]; Resulta[i,30]=s_p$li[1]; Resulta[i,31]=s_p$ls[1]
   Resulta[i,32]=s_p$ls[1]-s_p$li[1]
   Resulta[i,33]=Resulta[i,28]/Resulta[i,32]
   Resulta[i,34]=h_p$p[2]; Resulta[i,35]=h_p$li[2]; Resulta[i,36]=h_p$ls[2]
   Resulta[i,37]=h_p$ls[2]-h_p$li[2]
   Resulta[i,38]=s_p$p[2]; Resulta[i,39]=s_p$li[2]; Resulta[i,40]=s_p$ls[2]
   Resulta[i,41]=s_p$ls[2]-s_p$li[2]
   Resulta[i,42]=Resulta[i,37]/Resulta[i,41]
} ##### Fin de la simulación
##### Almacenando y dando forma a los resultados
TabResulta <- data.frame(Id=Resulta[,1], Dp=round(Resulta[,2],digits=1),
 h_b1=round(Resulta[,3],digits=3), h_b1_li=round(Resulta[,4],digits=3),
```

```
 h_b1_ls=round(Resulta[,5],digits=3), h_b1Dl=round(Resulta[,6],digits=3),
 h_b1Con=Resulta[,7], s_b1=round(Resulta[,8],digits=3), s_b1_li=round(Resulta[,9],digits=3),
 s_b1_ls=round(Resulta[,10],digits=3), s_b1Dl=round(Resulta[,11],digits=3),
 s_b1Con=Resulta[,12], b1Rl=round(Resulta[,13],digits=2),
 h_b2=round(Resulta[,14],digits=3), h_b2_li=round(Resulta[,15],digits=3),
 h_b2_ls=round(Resulta[,16],digits=3), h_b2Dl=round(Resulta[,17],digits=3),
 h_b2Con=Resulta[,18], s_b2=round(Resulta[,19],digits=3),
 s_b2_li=round(Resulta[,20],digits=3), s_b2_ls=round(Resulta[,21],digits=3),
 s_b2Dl=round(Resulta[,22],digits=3), s_b2Con=Resulta[,23],
 b2Rl=round(Resulta[,24],digits=2), h_p1=round(Resulta[,25],digits=3),
 h_p1_li=round(Resulta[,26],digits=3), h_p1_ls=round(Resulta[,27],digits=3),
 h_p1Dl=round(Resulta[,28],digits=3), s_p1=round(Resulta[,29],digits=3),
 s_p1_li=round(Resulta[,30],digits=3), s_p1_ls=round(Resulta[,31],digits=3),
 s_p1Dl=round(Resulta[,32],digits=3), p1Rl=round(Resulta[,33],digits=2),
 h_p2=round(Resulta[,34],digits=3), h_p2_li=round(Resulta[,35],digits=3),
 h_p2_ls=round(Resulta[,36],digits=3), h_p2Dl=round(Resulta[,37],digits=3),
 s_p2=round(Resulta[,38],digits=3), s_p2_li=round(Resulta[,39],digits=3),
 s_p2_ls=round(Resulta[,40],digits=3), s_p2Dl=round(Resulta[,41],digits=3),
 p2Rl=round(Resulta[,42],digits=2)   )
write.table(TabResulta, file = "C:/Resulta.txt",
      append = FALSE, dec = ".", quote = FALSE, sep = ";",
      row.names = FALSE, col.names = TRUE)
TabCon <- data.frame(Id=TabResulta$Id, Dp=TabResulta$Dp,
   Db1=round(abs(TabResulta$h_b1-TabResulta$s_b1), digits=3),
   b1Con=abs(TabResulta$h_b1Con-TabResulta$s_b1Con), b1Rl=TabResulta$b1Rl,
   Db2=round(abs(TabResulta$h_b2-TabResulta$s_b2), digits=3),
   b2Con=abs(TabResulta$h_b2Con-TabResulta$s_b2Con), b2Rl=TabResulta$b2Rl,
   Dp1=round(abs(TabResulta$h_p1-TabResulta$s_p1),digits=3), p1Rl=TabResulta$p1Rl,
   Dp2=round(abs(TabResulta$h_p2-TabResulta$s_p2),digits=3), p2Rl=TabResulta$p2Rl )
write.table(TabCon, file = "C:/Conclu.txt",
      append = FALSE, dec = ".", quote = FALSE, sep = ";",
      row.names = FALSE, col.names = TRUE)
##### Análisis posterior
with(TabCon, table(Dp, Db1)); with(TabCon, table(Dp, Db2))
with(TabCon, table(Dp, Dp1)); with(TabCon, table(Dp, Dp2))
b1Rl_Med <- with(TabCon, aggregate(b1Rl, by=list(Dp), FUN=mean))
b1Rl_SD  <- with(TabCon, aggregate(b1Rl, by=list(Dp), FUN=sd))
b2Rl_Med <- with(TabCon, aggregate(b2Rl, by=list(Dp), FUN=mean))
b2Rl_SD  <- with(TabCon, aggregate(b2Rl, by=list(Dp), FUN=sd))
p1Rl_Med <- with(TabCon, aggregate(p1Rl, by=list(Dp), FUN=mean))
p1Rl_SD  <- with(TabCon, aggregate(p1Rl, by=list(Dp), FUN=sd))
p2Rl_Med <- with(TabCon, aggregate(p2Rl, by=list(Dp), FUN=mean))
p2Rl_SD  <- with(TabCon, aggregate(p2Rl, by=list(Dp), FUN=sd))
TabRes1 <- data.frame(cbind(Dp=round(b1Rl_Med[,1],2), b1Rl_Med=round(b1Rl_Med[,2],2),
      b1Rl_SD=round(b1Rl_SD[,2],3),   b2Rl_Med=round(b2Rl_Med[,2],2),
      b2Rl_SD=round(b2Rl_SD[,2],3), p1Rl_Med=round(p1Rl_Med[,2],2),
      p1Rl_SD=round(p1Rl_SD[,2],3), p2Rl_Med=round(p2Rl_Med[,2],2),
      p2Rl_SD=round(p2Rl_SD[,2],3)) )
TabRes1
h_b1_0 <- with(Resulta, aggregate(1-h_b1Con, by=list(Dp), FUN=sum))
h_b1_1 <- with(Resulta, aggregate(h_b1Con, by=list(Dp), FUN=sum))
s_b1_0 <- with(Resulta, aggregate(1-s_b1Con, by=list(Dp), FUN=sum))
s_b1_1 <- with(Resulta, aggregate(s_b1Con, by=list(Dp), FUN=sum))
h_b2_0 <- with(Resulta, aggregate(1-h_b2Con, by=list(Dp), FUN=sum))
h_b2_1 <- with(Resulta, aggregate(h_b2Con, by=list(Dp), FUN=sum))
s_b2_0 <- with(Resulta, aggregate(1-s_b2Con, by=list(Dp), FUN=sum))
```

```
s_b2_1 <- with(Resulta, aggregate(s_b2Con, by=list(Dp), FUN=sum))
tCon <- data.frame(cbind(Dp=round(h_b1_0[,1],1),
      round(h_b1_0=h_b1_0[,2],0), round(h_b1_1=h_b1_1[,2],0),
      round(s_b1_0=s_b1_0[,2],0), round(s_b1_1=s_b1_1[,2],0),
      round(h_b2_0=h_b2_0[,2],0), round(h_b2_1=h_b2_1[,2],0),
      round(s_b2_0=s_b2_0[,2],0), round(s_b2_1=s_b2_1[,2],0)))
tCon
```