Chilean Journal of Statistics
Vol. 12, No. 1, April 2021, 83–102

Statistical Modeling
Research Paper

Alternatives to the logit model in the situation of factor levels aggregation in binomial responses

Malinda Coa[1] and Ernesto Ponsot[2,3,*]

[1]Institute of Applied Statistics and Computing, University of Los Andes, Mérida, Venezuela
[2]Statistics Department, University of Los Andes, Mérida, Venezuela
[3]Topos: Data Science Consulting, Ibarra, Ecuador

(Received: 27 October 2020 · Accepted in final form: 26 April 2021)

Abstract

The detailed study of the logit, probit and cloglog link functions is presented for the generalized linear model with binomial response in the presence of the problem of explanatory factors levels aggregation. Expressions are deduced for the estimators of the parameters and their variances, in general terms, which allows for finding the particular results for any link function chosen. The impact of the link function on the estimates is illustrated, concluding that the use of the appropriate variance in the levels aggregation is preferable, regardless of the link function to be used.

Keywords: Binomial regression · Generalized linear models · Level sets · Link function

Mathematics Subject Classification: Primary 62J12 · Secondary 62J20.

1. Introduction

The binomial model pursues the same objectives as the classical regression model, however, they differ in their structure (Collet, 2002; Tutz, 2011). The crucial difference is that, in the binomial model, the dependent variable follows a distribution that takes only two possible values, zero and one, in contrast to the normally distributed response of the classical regression model, in which any real value can be observed. Thus, the categorization of the response variable in the binomial model leads to the second difference. This concerns the need for a link function between the explanatory variables and the mean of the response.

Within the context of binomial models, the logit model is the most widely studied and applied (Christensen, 1997; Hilbe, 2009; Hosmer and Lemeshow, 2000). It is a particular case of the generalized linear model (Nelder and Wedderburn, 1972), when logit is the link function between the random component and the systematic component of the model, with the probability of success in a Bernoulli trial being modeled. Factors or treatments, rather than variables, are postulated in the style of the conventional analysis of variance (McCullagh and Nelder, 1989; McCulloch and Searle, 2001).

In its simplest formulation, the logit model consists of a dichotomous response variable and a single explanatory factor. Additionally, it is assumed that the responses corresponding

* Corresponding author. Email: malindacoa@gmail.com, ernesto.pb@gmail.com

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)
© Chilean Statistical Society – Sociedad Chilena de Estadística
http://www.soche.cl/chjs

to the different levels of the explanatory factor are independent binomials. After adjusting and applying this model on data in a contingency table, suppose that the researcher decides to group some levels of the factor and reiterate the logit analysis in the usual way. With this procedure, Ponsot et al. (2009) demonstrated that a violation of the binomial assumption is incurred, with important implications for the variance. They suggest courses of action to correct the problem and at the same time improve the accuracy of the results. Specifically, based on the reference parameterization and the saturated model, the authors suggest a procedure that takes advantage of the computations of a first logit adjustment and corrects the distributional assumption about variance, producing more efficient estimates and with greater precision than those obtained if you decide to reiterate a logit adjustment. Through simulations, strong trends were shown in favor of the proposed method, even more, if the probabilities of success of the response variable associated are more dissimilar to each other.

Note that the aforementioned research is limited to the scope of the logit model, but what about the problem of factor levels aggregation in models that are usually competitors or alternatives to the logit model? The logit model is the most used for the advantages it offers, however, it does not always guarantee a good fit for all binomial response data, so the researcher may consider other alternatives (Bonat et al., 2018; Czado and Santner, 1992; Czado and Munk, 2000). Logit is the canonical link function for binomial response data, but probit is also popular (McCullagh and Nelder, 1989; Collet, 2002; Hosmer and Lemeshow, 2000). In fact, any differentiable monotonous function can serve as a link between the random and systematic components of the binomial model, so there are many other functions that could offer a better fit than the traditional logit model. Therefore, keeping the problem within the scope of generalized linear models, this work seeks to answer this question by generalizing the procedure suggested by Ponsot et al. (2009) so that it is applicable with any link function.

The paper has been organized as follows: Section 2 explains the problem of explanatory factors levels aggregation. In Section 3, the fit of binomial models under this situation is addressed through the usual method. In Section 4, we describe the method suggested by Ponsot et al. (2009). Section 5 develops the adjustment procedures of three of the best known binomial models (logit, probit and cloglog), using the methods proposed in Sections 3 and 4. An example of the potential application of these models is shown in Section 6. Finally, Section 7 presents the main conclusions derived from this work.

2. The problem of explanatory factors levels aggregation

In Table 1, let y_i be the observed number of successes observed in the i-th level of the factor A and n_i the total number of observations for that level.

Table 1. Observed number of successes and total of the Y response versus the A factor levels.

	Y	
A	Number of successes	Total
1	y_1	n_1
2	y_2	n_2
$\vdots$	$\vdots$	$\vdots$
$k-2$	y_{k-2}	n_{k-2}
$k-1$	y_{k-1}	n_{k-1}
k	y_k	n_k
Total	$y_\cdot$	$n_\cdot$

The corresponding responses to the different levels of A are assumed independent of each other and from a binomial population in the number of successes ($Y = 1$). This is

$$Y_i \overset{\text{ind}}{\sim} \text{Bin}(y_i; n_i, p_i), \quad i = 1, \ldots, k$$

where "ind" stands for independent, Y_i is the random variable that represents the number of successes in the i-th sample and p_i, considered constant, is the probability of success associated ($0 < p_i < 1$). Assuming a binomial distribution in the number of successes of the Y_i at each level of the explanatory factor, implies that $\text{V}[Y_i] = n_i p_i(1-p_i)$ and $\text{E}[Y_i] = n_i p_i$.

Sometimes, after adjusting a model the researcher may decide to group levels of the A factor for various reasons. Suppose the levels k and $k-1$ are added doing $y^*_{k-1} = y_{k-1} + y_k$ and $n^*_{k-1} = n_{k-1} + n_k$, obtaining an arrangement of the data as in Table 2. The situation could extend to more than two levels, simply by adding the last two, then these with the previous one, and so on.

Table 2. Number of successes and total of the Y response versus the A factor levels, after the aggregation of the k and $k-1$ levels.

	Y	
A	Number of successes	Total
1	y_1	n_1
2	y_2	n_2
$\vdots$	$\vdots$	$\vdots$
$k-2$	y_{k-2}	n_{k-2}
$k-1$	y^*_{k-1}	n^*_{k-1}
Total	$y_\cdot$	$n_\cdot$

Usually, by reiterating the model fit procedure, the researcher assumes that the new random variable $Y^*_{k-1} = Y_{k-1} + Y_k$, that arises from aggregation, still has a binomial distribution with variance

$$\text{V}_{\text{Bin}}[Y^*_{k-1}] = n^*_{k-1} p^*_{k-1}(1 - p^*_{k-1}), \tag{1}$$

where $n^*_{k-1} = n_{k-1} + n_k$ and $p^*_{k-1} = \text{E}[Y^*_{k-1}]/n^*_{k-1} = (n_{k-1}p_{k-1} + n_k p_k)/(n_{k-1} + n_k)$.

In this regard, Ponsot et al. (2009) demonstrated that with this proceeding, a violation of the original binomial assumption is incurred, with important implications for the estimated variances. In their work, the authors deduced the following:

(1) If $p_{k-1} \neq p_k$, the binomial assumption is violated in the sample corresponding to the level of the response variable where aggregation arises: Y^*_{k-1} is actually distributed Poisson-binomial (and not binomial).
(2) The right expression for the variance of Y^*_{k-1} is

$$\text{V}[Y^*_{k-1}] = n_{k-1}p_{k-1}(1 - p_{k-1}) + n_k p_k(1 - p_k). \tag{2}$$

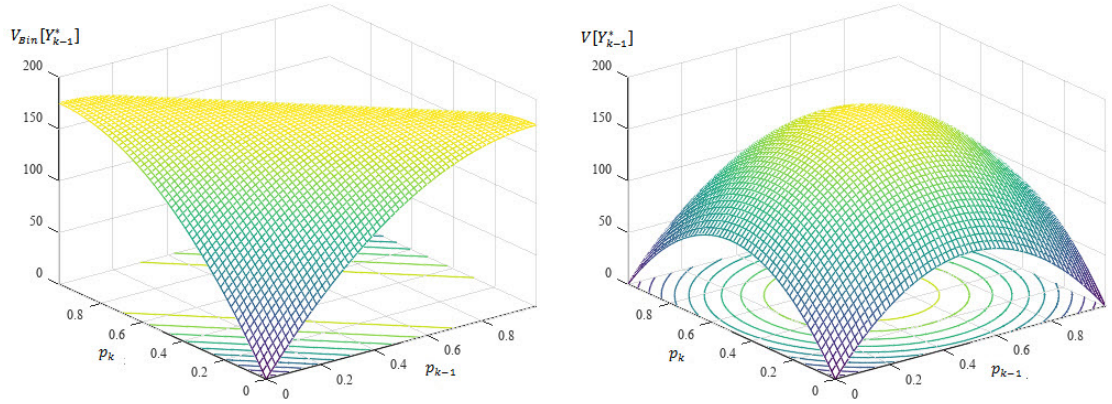
Also, the authors argued that Equations (1) and (2) are not equivalent and that in general it has $\text{V}_{\text{Bin}}[Y^*_{k-1}] \geq \text{V}[Y^*_{k-1}]$. In this regard, let $\Delta V = \text{V}_{\text{Bin}}[Y^*_{k-1}] - V[Y^*_{k-1}]$. Then, we have

$$\Delta V = \frac{n_{k-1}n_k}{n_{k-1} + n_k}(p_{k-1} - p_k)^2. \tag{3}$$

From Equation 3, the authors deduced that:

(1) If p_{k-1} and p_k are close ($p_{k-1} \approx p_k$), the difference $(p_{k-1}-p_k)^2 \to 0$, so that $\Delta V \to 0$ and $\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}] \approx \mathrm{V}[Y^*_{k-1}]$.
(2) If p_{k-1} and p_k are distant from each other (which occurs, for example, when $p_{k-1} \approx 0$ and $p_k \approx 1$, or $p_{k-1} \approx 1$ and $p_k \approx 0$), then the difference $(p_{k-1} - p_k)^2 \to 1$, so that $\Delta V \to n_{k-1}n_k/(n_{k-1}+n_k)$ and $\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}] \approx \mathrm{V}[Y^*_{k-1}] + n_{k-1}n_k/(n_{k-1}+n_k)$. In this situation, the greatest difference between the variances occurs. This difference can be considerable depending on the n_{k-1} and n_k values.

Now, Figure 1(a) shows the behavior of the variance assumed by the researcher when the n_{k-1} and n_k parameters are fixed and vary the values of p_{k-1} and p_k. Clearly, it can be seen in the graph that as $p_{k-1} \approx 0$ and $p_k \approx 0$, or $p_{k-1} \approx 1$ and $p_k \approx 1$, then $\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}] \to 0$. Meanwhile, relative maximums are obtained along the ordered pairs $(p_{k-1}, p_k) = (p_{k-1}, 0.5[1+(n_{k-1}/n_k)(1-2p_{k-1})])$, where $\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]$ reaches the value $n^*_{k-1}/4$ in each one of them (see Appendix A).



(a) (b)

Figure 1. Y^*_{k-1} variances: (a) Binomial variance ($\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]$) assumed by the researcher; (b) True variance ($\mathrm{V}[Y^*_{k-1}]$).

Note that the true variance (Figure 1(b)) shows similar behavior to the binomial variance, when p_{k-1} and p_k both tend to 0 or 1. In fact, whenever $p_{k-1} = p_k$. However, this behavior also occurs when one of them tends to 0 and the other to 1 (or viceversa). The minimum values of the variance occur for the cases mentioned, while the maximum occurs when both parameters take the value $0,5$ ($p_{k-1} = p_k = 0.5$), being $\mathrm{V}[Y^*_{k-1}] = n^*_{k-1}/4$ said maximum (see Appendix B).

In Figure 2, the difference between the variance assumed by the researcher and the true variance is shown. Indeed, this figure shows that the differences tend to 0 when $p_{k-1} \approx p_k$. These differences grow when p_{k-1} and p_k tend to opposite ends, reaching their maximum values in $p_{k-1} \approx 0$ and $p_k \approx 1$, and $p_{k-1} \approx 1$ and $p_k \approx 0$.

3. Fitting a binomial model using the usual method

Next, we describe the saturated binomial model before aggregation. It is possible to use the generalized linear model approach (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989; McCulloch and Searle, 2001; Dobson, 2002; Agresti, 2007, 2015) to the data in Table 1, when the response variables $Y_1, \ldots, Y_k$ are supposed independent and follow a binomial

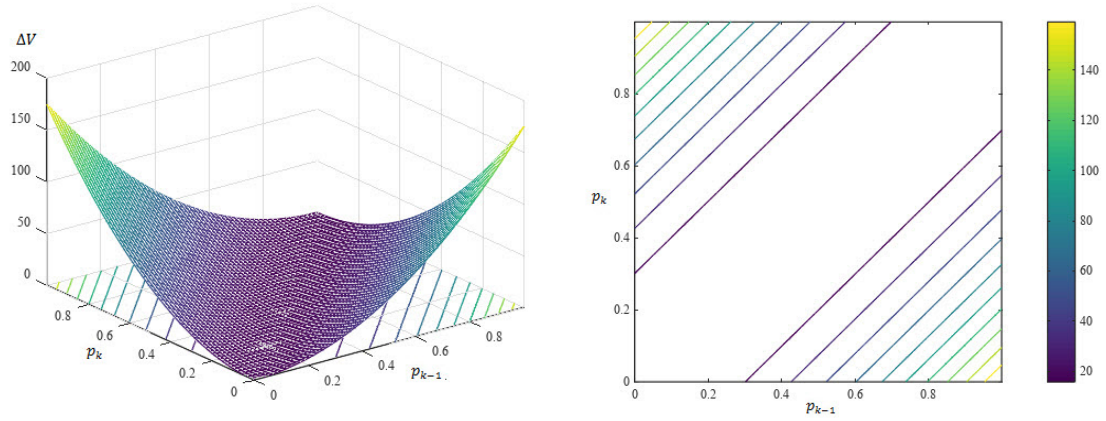


Figure 2. Difference between binomial and true variances ($\Delta V = \mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}] - \mathrm{V}[Y^*_{k-1}]$).

distribution as this belongs to the exponential family of distributions. For the link, any monotonous and differentiable function can be used, however, the choice of it has given rise to the most important binomial models present in the literature (logit model, probit model, cloglog model, to mention some, among which the first one stands out).

In its simplest sense, the saturated model ($m = k$) and reference parameterization can be postulated, whereby the matrix $\boldsymbol{X}$ is a square matrix (of order $k \times k$) and invertible (Ponsot, 2011). Being k the reference level, this parameterization leads to the model $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$, whose matrix representation is:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{k-2} \\ \eta_{k-1} \\ \eta_k \end{bmatrix} = \begin{bmatrix} 1\,1\,0\,0\cdots 0\,0 \\ 1\,0\,1\,0\cdots 0\,0 \\ \vdots\,\vdots\,\vdots\,\vdots\,\ddots\,\vdots\,\vdots \\ 1\,0\,0\,0\cdots 1\,0 \\ 1\,0\,0\,0\cdots 0\,1 \\ 1\,0\,0\,0\cdots 0\,0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-2} \\ \beta_{k-1} \\ \beta_k \end{bmatrix}. \tag{4}$$

The $\boldsymbol{X}$ matrix can be partitioned as:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{j} & \boldsymbol{I} \\ 1 & \boldsymbol{0}^\top \end{bmatrix} \Rightarrow \boldsymbol{X}^{-1} = \begin{bmatrix} \boldsymbol{0}^\top & 1 \\ \boldsymbol{I} & -\boldsymbol{j} \end{bmatrix}. \tag{5}$$

Since $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} \Rightarrow \boldsymbol{\beta} = \boldsymbol{X}^{-1}\boldsymbol{\eta}$. When saturated, the model raised in Equation (4) does not have sufficient degrees of freedom to calculate the *deviance* or Pearson statistics (Ponsot, 2011). However, you can still estimate its parameters ($\boldsymbol{\beta}$) and determine its statistical significance.

Let $\widehat{\eta}_i = g(\widehat{\mu}_i), i = 1, \ldots, k$. From Equations (4) and (5), it follows in general terms that

$$\widehat{\beta}_j = \begin{cases} \widehat{\eta}_k & \text{if } j = 1 \\ \widehat{\eta}_{j-1} - \widehat{\eta}_k & \text{if } j = 2, \ldots, k. \end{cases} \tag{6}$$

Since $\boldsymbol{\beta} \sim \mathrm{AN}[\boldsymbol{\beta}, (\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X})^{-1}]$, the parameters variance is

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}] = (\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X})^{-1} = \boldsymbol{X}^{-1}\boldsymbol{W}^{-1}(\boldsymbol{X}^\top)^{-1}$$

where $\boldsymbol{W} = \mathrm{diag}\{w_1, \ldots, w_k\}$ with $w_i = (\partial \mu_i / \partial \eta_i)^2 \mathrm{V}_{\mathrm{Bin}}[Y_i]$, for $i = 1, \ldots, k$.

Therefore, in general terms, the variance of $\widehat{\boldsymbol{\beta}}$ is given by

$$\mathrm{V}[\widehat{\beta}_j] = \begin{cases} \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_k]}{(\partial\mu_k/\partial\eta_k)^2}, & \text{if } j = 1; \\ \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]}{(\partial\mu_{j-1}/\partial\eta_{j-1})^2} + \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_k]}{(\partial\mu_k/\partial\eta_k)^2}, & \text{if } j = 2, \ldots, k. \end{cases} \tag{7}$$

After fitting a saturated binomial model, assume that the last two levels $k-1$ and k of the A factor are added, leaving the data arranged as in Table 2. As already mentioned sometimes the researcher reiterates the usual adjustment process, as in the previous section, assuming that the new random variable $Y^*_{k-1} = Y_{k-1} + Y_k$ has variance $\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}] = n^*_{k-1}p^*_{k-1}(1 - p^*_{k-1})$, with $n^*_{k-1} = n_{k-1} + n_k$ and $p^*_{k-1} = (n_{k-1}p_{k-1} + n_kp_k)/n^*_{k-1}$.

In this new fit, the design matrix that arises from aggregation (call $\boldsymbol{X}^*$) now has dimensions $(k-1)\times(k-1)$ due to the elimination of the k-th row and k-th column. However, if the reference parameterization is maintained, then with respect to the aggregate level $k-1$ is possible to propose a model as in Equation (4) on the new data set.

As in Equation (6), the estimated parameter vector elements obtained by the usual method (denote the superscript h like $\widehat{\boldsymbol{\beta}}^{*h}$) are expressed as

$$\widehat{\beta}^{*h}_j = \begin{cases} \widehat{\eta}^*_{k-1}, & \text{if } j = 1; \\ \widehat{\eta}_{j-1} - \widehat{\eta}^*_{k-1}, & \text{if } j = 2, \ldots, k-1. \end{cases} \tag{8}$$

Observe that the new diagonal matrix $\boldsymbol{W}^{*h}$ keeps the $k-2$ elements of $\boldsymbol{W}$, changing only the one that corresponds to the level $k-1$, that is, we have

$$\begin{aligned} \boldsymbol{W}^{*h} &= \mathrm{diag}\left\{w_1, \ldots, w^*_{k-1}\right\} \\ &= \mathrm{diag}\left\{\frac{(\partial\mu_1/\partial\eta_1)^2}{\mathrm{V}_{\mathrm{Bin}}[Y_1]}, \ldots, \frac{(\partial\mu^*_{k-1}/\partial\eta^*_{k-1})^2}{\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]}\right\}. \end{aligned}$$

Thus, the variances and covariances matrix is given by

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}^{*h}] = [(\boldsymbol{X}^*)^\top\boldsymbol{W}^{*h}\boldsymbol{X}^*]^{-1} = (\boldsymbol{X}^*)^{-1}(\boldsymbol{W}^{*h})^{-1}[(\boldsymbol{X}^*)^\top]^{-1}$$

and as in Equation (7), the variance of $\widehat{\beta}^{*h}_j$ is

$$\mathrm{V}[\widehat{\beta}^{*h}_j] = \begin{cases} \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]}{\left(\partial\mu^*_{k-1}/\partial\eta^*_{k-1}\right)^2}, & \text{if } j = 1; \\ \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]}{(\partial\mu_{j-1}/\partial\eta_{j-1})^2} + \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]}{\left(\partial\mu^*_{k-1}/\partial\eta^*_{k-1}\right)^2}, & \text{if } j = 2, \ldots, k-1. \end{cases} \tag{9}$$

Equations (8) and (9) are generalizations of the usual procedure for fitting binomial models when any differentiable monotone function is used as a link. Therefore, from now on, they are part of the context that we call the generalized usual method (GUM).

4. Fitting a binomial model using the Ponsot method

In the presence of the saturated model given in Equation (4), Ponsot et al. (2009) proposed a method to reiterate the adjustment of a binomial model, after the levels aggregation, when the first adjustment uses the logit link function.

Now, let $g(\mu_i)$ be any differentiable monotonous link function. Asymptotically, in the saturated model $\mathrm{V}[\boldsymbol{X}\widehat{\boldsymbol{\beta}}] = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{X}\boldsymbol{X}^{-1}\boldsymbol{W}^{-1}(\boldsymbol{X}^\top)^{-1}\boldsymbol{X}^\top = \boldsymbol{W}^{-1}$. Then, under conditions of regularity and as a consequence of the usual central limit theorem, as well as properties of the maximum-likelihood estimators, we have that

$$g(\widehat{\mu}_i) = x_i^\top \widehat{\boldsymbol{\beta}} \sim \mathrm{AN}\left(x_i^\top \boldsymbol{\beta}; w_i^{-1} = \left[\frac{(\partial\mu_i/\partial\eta_i)^2}{\mathrm{V}_{\mathrm{Bin}}[Y_i]}\right]^{-1} = \frac{\mathrm{V}_{\mathrm{Bin}}[Y_i]}{(\partial\mu_i/\partial\eta_i)^2}\right). \tag{10}$$

Applying the delta method in Equation (10) (Agresti, 2007), we get

$$\begin{aligned} \mu_i &= g^{-1}(\eta_i) \\ \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} &= \frac{\partial \mu_i}{\partial \eta_i}. \end{aligned}$$

Thus, we reach

$$\widehat{\mu}_i \sim \mathrm{AN}\left(\mu_i = g^{-1}(x_i^\top \boldsymbol{\beta}); \mathrm{V}_{\mathrm{Bin}}[Y_i]\right).$$

Now, when $k-1$ and k levels are added, the maximum-likelihood estimator of the new mean μ^*_{k-1} is given by

$$\widehat{\mu}^*_{k-1} = \widehat{\mathrm{E}[Y^*_{k-1}]} = \widehat{\mathrm{E}[Y_{k-1}]} + \widehat{\mathrm{E}[Y_k]} = \widehat{\mu}_{k-1} + \widehat{\mu}_k. \tag{11}$$

Due to $\widehat{\mu}^*_{k-1}$ is the weighted sum of two linear functions of asymptotically independent normal random variables, their asymptotic distribution is also normal with

$$\mathrm{E}[\widehat{\mu}^*_{k-1}] = \mathrm{E}[\widehat{\mu}_{k-1}] + \mathrm{E}[\widehat{\mu}_k] = \mu_{k-1} + \mu_k$$

and

$$\mathrm{V}[\widehat{\mu}^*_{k-1}] = \mathrm{V}[\widehat{\mu}_{k-1}] + \mathrm{V}[\widehat{\mu}_k] = \mathrm{V}_{\mathrm{Bin}}[Y_{k-1}] + \mathrm{V}_{\mathrm{Bin}}[Y_k] = \mathrm{V}[Y^*_{k-1}].$$

Again, using the delta method, the required distribution of $\widehat{\eta}^*_{k-1} = g(\widehat{\mu}^*_{k-1})$ is asymptotically normal with expected value $\mathrm{E}[\widehat{\eta}^*_{k-1}] = g(\mu^*_{k-1}) = \eta^*_{k-1}$ and asymptotic variance stated as

$$(\sigma^2)^*_{k-1} = \mathrm{V}[\widehat{\mu}^*_{k-1}]\left[\frac{\partial g(\mu^*_{k-1})}{\partial \mu^*_{k-1}}\right]^2 = \mathrm{V}[Y^*_{k-1}]\left[\frac{\partial \eta^*_{k-1}}{\partial \mu^*_{k-1}}\right]^2. \tag{12}$$

From $(\sigma^2)^*_{k-1}$, Ponsot et al. (2009) suggested creating a matrix $\boldsymbol{\Sigma}$, equivalent to $\boldsymbol{W}^{-1}$ from the original fit and from which its $k-2$ elements remains, but with the $k-1$ element added corrected for true variance $\mathrm{V}[Y^*_{k-1}]$, as in Equation (12). This suggested matrix is of

the form expressed as

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1/w_1 & 0 & \cdots & 0 & 0 \\ 0 & 1/w_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1/w_{k-2} & 0 \\ 0 & 0 & \cdots & 0 & (\sigma^2)^*_{k-1} \end{bmatrix}.$$

The variance and covariance matrix of the estimators using this suggested method (now denoted with the superscript s) is given by

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}^{*s}] = (\boldsymbol{X}^*)^{-1}\boldsymbol{\Sigma}[(\boldsymbol{X}^*)^\top]^{-1}$$

resulting in

$$\mathrm{V}[\widehat{\beta}^{*s}_j] = \begin{cases} \dfrac{\mathrm{V}[Y^*_{k-1}]}{\left(\partial\mu^*_{k-1}/\partial\eta^*_{k-1}\right)^2}, & \text{if } j = 1; \\ \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]}{\left(\partial\mu_{j-1}/\partial\eta_{j-1}\right)^2} + \dfrac{\mathrm{V}[Y^*_{k-1}]}{\left(\partial\mu^*_{k-1}/\partial\eta^*_{k-1}\right)^2}, & \text{if } j = 2, \cdot, k-1. \end{cases} \tag{13}$$

As in Equation (8), the elements of the estimated parameter vector using the suggested method ($\widehat{\boldsymbol{\beta}}^{*s}$) are defined as

$$\widehat{\beta}^{*s}_j = \begin{cases} \widehat{\eta}^*_{k-1} & \text{if } j = 1 \\ \widehat{\eta}_{j-1} - \widehat{\eta}^*_{k-1} & \text{if } j = 2, \cdots, k-1. \end{cases} \tag{14}$$

The above results are a generalization of the method suggested by Ponsot et al. (2009), which we call the generalized suggested method (GSM). The GSM allows us to adjust a binomial model using the suggested method, but now using any link function.

Next, we describe differences between the variances obtained through GUM and GSM. Let $\Delta\mathrm{V}(\widehat{\beta}^*_1) = \mathrm{V}[\beta^{*h}_1] - \mathrm{V}[\beta^{*s}_1]$ be the difference between the variances obtained by the GUM and the GSM, for the β^*_1 parameter. Then, from Equations (9) and (13), we have

$$\begin{aligned} \Delta\mathrm{V}(\widehat{\beta}^*_1) &= \frac{\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]}{\left(\partial\mu^*_{k-1}/\partial\eta^*_{k-1}\right)^2} - \frac{\mathrm{V}[Y^*_{k-1}]}{\left(\partial\mu^*_{k-1}/\partial\eta^*_{k-1}\right)^2}, \\ &= \frac{\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}] - \mathrm{V}[Y^*_{k-1}]}{\left(\partial\mu^*_{k-1}/\partial\eta^*_{k-1}\right)^2}, \\ &= \frac{\Delta\mathrm{V}}{\left(\partial\mu^*_{k-1}/\partial\eta^*_{k-1}\right)^2}, \end{aligned}$$

where $\Delta\mathrm{V}$ is defined as in Equation (3).

For the parameters β_j^*, with $j = 2, \ldots, , k-1$, let $\Delta \mathrm{V}(\widehat{\beta}_j^*) = \mathrm{V}(\widehat{\beta}_j^{*h}) - \mathrm{V}(\widehat{\beta}_j^{*s})$ be the difference between the variances obtained by both methods. Again, from Equations (9) and (13), we get

$$\begin{aligned}
\Delta \mathrm{V}(\widehat{\beta}_j^*) &= \left[\frac{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]}{(\partial \mu_{j-1}/\partial \eta_{j-1})^2} + \frac{\mathrm{V}_{\mathrm{Bin}}[Y_{k-1}^*]}{\left(\partial \mu_{k-1}^*/\partial \eta_{k-1}^*\right)^2} \right] \\
&\quad - \left[\frac{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]}{(\partial \mu_{j-1}/\partial \eta_{j-1})^2} + \frac{\mathrm{V}[Y_{k-1}^*]}{\left(\partial \mu_{k-1}^*/\partial \eta_{k-1}^*\right)^2} \right] \\
&= \frac{\mathrm{V}_{\mathrm{Bin}}[Y_{k-1}^*]}{\left(\partial \mu_{k-1}^*/\partial \eta_{k-1}^*\right)^2} - \frac{\mathrm{V}[Y_{k-1}^*]}{\left(\partial \mu_{k-1}^*/\partial \eta_{k-1}^*\right)^2} \\
&= \mathrm{V}[\beta_1^{*h}] - \mathrm{V}[\beta_1^{*s}] = \Delta \mathrm{V}(\widehat{\beta}_1^*).
\end{aligned}$$

This last result demonstrates that the difference in the variances obtained by both methods, for the first component of the vector $\widehat{\boldsymbol{\beta}}^*$ (that is, $\widehat{\beta}_1^*$), is the same for the remaining components. Thus, we have that

$$\Delta \mathrm{V}(\widehat{\beta}_j^*) = \frac{\Delta \mathrm{V}}{\left(\partial \mu_{k-1}^*/\partial \eta_{k-1}^*\right)^2}, \quad \text{for} \quad j = 1, \ldots, k-1. \tag{15}$$

In Equation (15), the denominator is dependent on the selected link function and the numerator is a function of $\Delta \mathrm{V}$. The same conditions established at the end of the Section 2 apply to p_{k-1} and p_k.

5. Some binomial models using the GUM and GSM

Assume the k and $k-1$ level aggregation situation in Table 2, and let

$$\widehat{p}_{k-1}^* = \frac{n_{k-1}\widehat{p}_{k-1} + n_k \widehat{p}_k}{n_{k-1} + n_k}, \tag{16}$$

as it follows from Equation (11). In this section, the adjustment procedure of three of the best known binomial models (logit, probit and cloglog) is shown, using the methods obtained in the previous sections.

The link function in the logit model is stated as

$$\eta_i = \mathrm{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \log\left(\frac{\mu_i}{n-\mu_i}\right).$$

Then, we have

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{n_i p_i (1-p_i)} = \frac{1}{\mathrm{V}_{\mathrm{Bin}}[Y_i]}.$$

Thus, we get

$$\frac{\partial \mu_i}{\partial \eta_i} = \mathrm{V}_{\mathrm{Bin}}[Y_i].$$

From Equations (8) and (14), the parameters estimated by both methods are identical and equal to

$$\widehat{\beta}_j^{*h} = \widehat{\beta}_j^{*s} = \begin{cases} \mathrm{logit}(\widehat{p}_{k-1}^{*}), & \text{if } j = 1; \\ \mathrm{logit}(\widehat{p}_{j-1}) - \mathrm{logit}(\widehat{p}_{k-1}^{*}), & \text{if } j = 2, \ldots, k-1. \end{cases}$$

The variances of the estimators obtained by GUM, according to Equation (9), are obtained as

$$\mathrm{V}[\widehat{\beta}_j^{*h}] = \begin{cases} \dfrac{1}{\mathrm{V}_{\mathrm{Bin}}[Y_{k-1}^{*}]}, & \text{if } j = 1; \\ \dfrac{1}{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]} + \dfrac{1}{\mathrm{V}_{\mathrm{Bin}}[Y_{k-1}^{*}]}, & \text{if } j = 2, \ldots, k-1. \end{cases}$$

Meanwhile, the variances estimated by the GSM for these same estimators, according to Equation (13), are given by

$$\mathrm{V}[\widehat{\beta}_j^{*s}] = \begin{cases} \dfrac{\mathrm{V}[Y_{k-1}^{*}]}{(\mathrm{V}_{\mathrm{Bin}}[Y_{k-1}^{*}])^2}, & \text{if } j = 1; \\ \dfrac{1}{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]} + \dfrac{\mathrm{V}[Y_{k-1}^{*}]}{(\mathrm{V}_{\mathrm{Bin}}[Y_{k-1}^{*}])^2}, & \text{if } j = 2, \ldots, k-1. \end{cases}$$

Regarding the differences between these variances, from Equation (15), we have that

$$\Delta\mathrm{V}(\widehat{\beta}_j^{*}) = \Delta\mathrm{V}(\widehat{\beta}_1^{*}) = \frac{\Delta\mathrm{V}}{(\mathrm{V}_{\mathrm{Bin}}[Y_{k-1}^{*}])^2}.$$

Figure 3 graphically displays the behavior of $\Delta\mathrm{V}(\widehat{\beta}_j^{*})$ for the logit model. It is observed that its performance is very similar to that of $\Delta\mathrm{V}$ shown in Figure 2.

The link function in the probit model is given by

$$\eta_i = \mathrm{probit}(p_i) = \Phi^{-1}(p_i) = \Phi^{-1}\left(\frac{n_i p_i}{n_i}\right) = \Phi^{-1}\left(\frac{\mu_i}{n_i}\right).$$

Hence, we get

$$\mu_i = n_i \Phi(\eta_i)$$

and since $[\Phi]' = \phi$, then we have that

$$\frac{\partial \mu_i}{\partial \eta_i} = n_i \phi(\eta_i) = n_i \phi[\Phi^{-1}(p_i)].$$

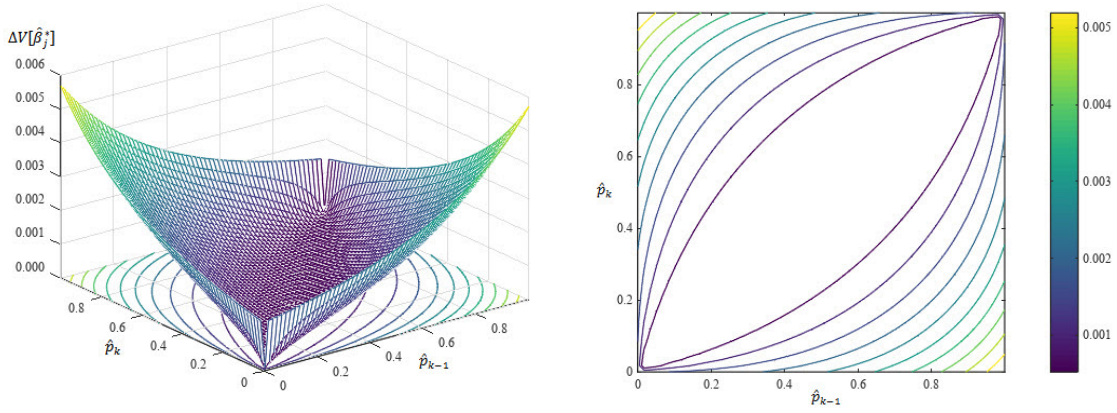


Figure 3. Differences between variances estimated using the GUM and GSM ($\Delta \mathrm{V}(\widehat{\beta}_j^*) = \mathrm{V}(\widehat{\beta}_j^{*h}) - V(\widehat{\beta}_j^{*s})$) with logit link function

From Equation (8) and (14), the parameters estimated by the GUM and GSM are identical and equal to

$$\widehat{\beta}_j^{*h} = \widehat{\beta}_j^{*s} = \begin{cases} \text{probit}(\widehat{p}_{k-1}^*), & \text{if } j = 1 \\ \text{probit}(\widehat{p}_{j-1}) - \text{probit}(\widehat{p}_{k-1}^*), & \text{if } j = 2, \ldots, k-1. \end{cases}$$

From Equation (9), the variances of the estimators obtained by the GUM are stated as

$$\mathrm{V}[\widehat{\beta}_j^{*h}] = \begin{cases} \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_{k-1}^*]}{\left\{n_{k-1}^* \phi[\Phi^{-1}(p_{k-1}^*)]\right\}^2}, & \text{if } j = 1; \\ \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]}{\{n_{j-1}\phi[\Phi^{-1}(p_{j-1})]\}^2} + \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_{k-1}^*]}{\left\{n_{k-1}^* \phi[\Phi^{-1}(p_{k-1}^*)]\right\}^2}, & \text{if } j = 2, \ldots, k-1. \end{cases}$$

Meanwhile, the variances estimated by the GSM, according to Equation (13), are expressed as

$$\mathrm{V}[\widehat{\beta}_j^{*s}] = \begin{cases} \dfrac{\mathrm{V}[Y_{k-1}^*]}{\left\{n_{k-1}^* \phi[\Phi^{-1}(p_{k-1}^*)]\right\}^2}, & \text{if } j = 1; \\ \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]}{\{n_{j-1}\phi[\Phi^{-1}(p_{j-1})]\}^2} + \dfrac{\mathrm{V}[Y_{k-1}^*]}{\left\{n_{k-1}^* \phi[\Phi^{-1}(p_{k-1}^*)]\right\}^2}, & \text{if } j = 2, \ldots, k-1. \end{cases}$$

From Equation (15), the differences between these estimated variances are defined by

$$\Delta \mathrm{V}(\widehat{\beta}_j^*) = \Delta \mathrm{V}(\widehat{\beta}_1^*) = \frac{\Delta \mathrm{V}}{\left\{n_{k-1}^* \phi[\Phi^{-1}(p_{k-1}^*)]\right\}^2}.$$

Figure 4 graphically displays the behavior of $\Delta \mathrm{V}(\widehat{\beta}_j^*)$ for the probit model. Without considering the difference of scales inherent in each case, it is observed that it is similar to that of the logit model, differing mainly in the borders or neighborhoods where $\widehat{p}_{k-1}$ and $\widehat{p}_k$ both approach 0 or 1.

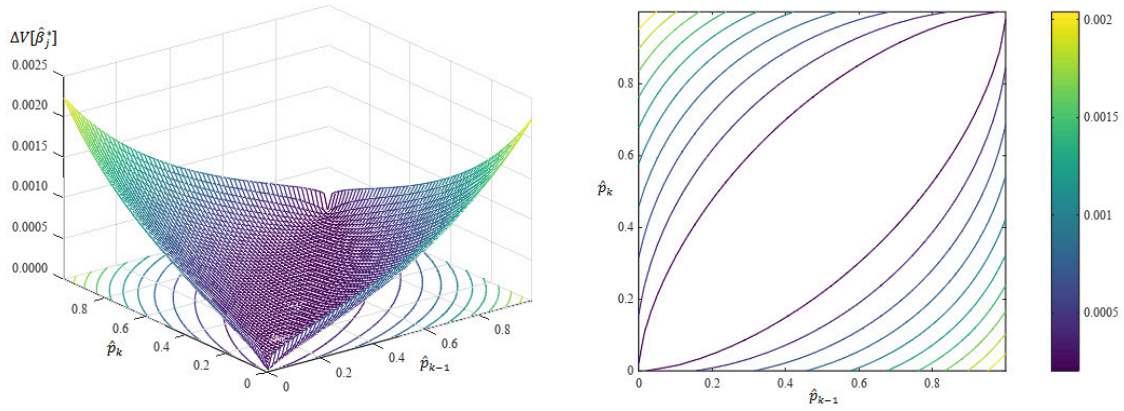


Figure 4. Differences between variances estimated using the GUM and GSM ($\Delta\mathrm{V}(\widehat{\beta}^*_j) = \mathrm{V}[\widehat{\beta}^{*h}_j] - \mathrm{V}[\widehat{\beta}^{*s}_j]$) with probit link function.

The link function in the cloglog model is established by

$$\eta_i = \mathrm{cloglog}(p_i) = \log[-\log(1-p_i)] = \log[-\log(1-\mu_i/n_i)] = g(\mu_i).$$

By differentiating, we obtain

$$\frac{\partial \eta_i}{\partial \mu_i} = -\frac{1}{n_i(1-p_i)\log(1-p_i)}.$$

Thus, we get

$$\frac{\partial \mu_i}{\partial \eta_i} = -n_i(1-p_i)\log(1-p_i).$$

From Equation (8) and (14), the parameters estimated by GUM and GSM are, once again, identical and equal to

$$\widehat{\beta}^{*h}_j = \widehat{\beta}^{*s}_j = \begin{cases} \mathrm{cloglog}(\widehat{p}^*_{k-1}), & \text{if } j = 1; \\ \mathrm{cloglog}(\widehat{p}_{j-1}) - \mathrm{cloglog}(\widehat{p}^*_{k-1}), & \text{if } j = 2, \ldots, k-1. \end{cases}$$

The variances of the estimators obtained by the GUM, according to Equation (9), are stated as

$$\mathrm{V}[\widehat{\beta}^{*h}_j] = \begin{cases} \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]}{[n^*_{k-1}(1-p^*_{k-1})\log(1-p^*_{k-1})]^2}, & \text{if } j = 1 \\ \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]}{[n_{j-1}(1-p_{j-1})\log(1-p_{j-1})]^2} & \\ + \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]}{[n^*_{k-1}(1-p^*_{k-1})\log(1-p^*_{k-1})]^2}, & \text{if } j = 2, \ldots, k-1. \end{cases}$$

Meanwhile, the variances estimated by the GSM for these same parameters, according to Equation (13), are given by

$$\mathrm{V}[\widehat{\beta}_j^{*s}] = \begin{cases} \dfrac{\mathrm{V}[Y_{k-1}^*]}{[n_{k-1}^*(1-p_{k-1}^*)\log(1-p_{k-1}^*)]^2}, & \text{if } j=1; \\ \dfrac{\mathrm{V}_{\mathrm{Bin}}[Y_{j-1}]}{[n_{j-1}(1-p_{j-1})\log(1-p_{j-1})]^2} & \\ +\dfrac{\mathrm{V}[Y_{k-1}^*]}{[n_{k-1}^*(1-p_{k-1}^*)\log(1-p_{k-1}^*)]^2}, & \text{if } j=2,\ldots,k-1. \end{cases}$$

From Equation (15), we get

$$\Delta\mathrm{V}(\widehat{\beta}_j^*) = \Delta\mathrm{V}(\widehat{\beta}_1^*) = \frac{\Delta\mathrm{V}}{[n_{k-1}^*(1-p_{k-1}^*)\log(1-p_{k-1}^*)]^2}.$$

Figure 5 shows the performance of $\Delta\mathrm{V}(\widehat{\beta}_j^*)$ for the cloglog model. Of course, it also constitutes a particularization of $\Delta\mathrm{V}$. However, it is observed that the region or border where $\Delta\mathrm{V}(\widehat{\beta}_j^*) \to 0$ is a little more extensive than the previous models.

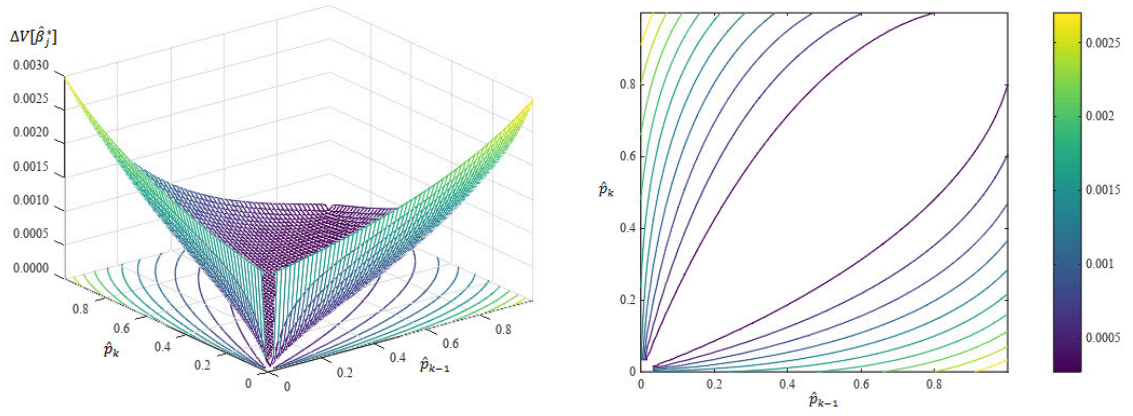


Figure 5. Differences between variances estimated using the GUM and GSM ($\Delta\mathrm{V}(\widehat{\beta}_j^*) = \mathrm{V}[\widehat{\beta}_j^{*h}] - \mathrm{V}[\widehat{\beta}_j^{*s}]$) with cloglog link function.

6. Illustration of the procedures

Table 3 reproduces the example presented by Ponsot (2011). There, the situation of interest focused on studying the relationship between a Y response variable and an A explanatory factor with three levels.

Table 3. Example: $Y(0,1)$ versus $A(1,2,3)$.

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

From Equation (4), it follows the saturated model using the parameterization with the third level of the factor as reference stated as

$$\begin{bmatrix}\eta_1\\ \eta_2\\ \eta_3\end{bmatrix} = \begin{bmatrix}g(p_1)\\ g(p_2)\\ g(p_3)\end{bmatrix} = \begin{bmatrix}1\ 1\ 0\\ 1\ 0\ 1\\ 1\ 0\ 0\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix}.$$

Table 4 contains the estimates of the parameters of the linear predictor and its variances for the different binomial models, according to the link function used (for this case, logit, probit and cloglog links). This Table also contains the Wald χ^2 tests for H_0: $\beta_i = 0$ in order to check the statistical significance of the estimated parameters, as well as the 95% confidence intervals (CI) built for β_i. The predicted probabilities and their CI, following Agresti (2007), are shown in Table 5.

Table 4. Original model: $\widehat{\beta}_i$ and Wald test (H_0: $\beta_i = 0$) according to the link function.

		Estimation of β_i					95% CI	
Link	i	$\widehat{\beta}_i$	$V[\widehat{\beta}_i]$	χ^2	p-value	Decision	LL	UL
logit	1	2.296	0.034	153.3	< 0.0001	Reject	1.933	2.660
	2	−2.457	0.046	131.5	< 0.0001	Reject	−2.877	−2.037
	3	−4.088	0.058	289.5	< 0.0001	Reject	−4.559	−3.617
probit	1	1.132	0.009	201.8	< 0.0001	Reject	1.148	1.516
	2	−1.432	0.013	154.3	< 0.0001	Reject	−1.658	−1.206
	3	−2.400	0.016	367.6	< 0.0001	Reject	−2.645	−2.154
cloglog	1	0.872	0.005	153.3	< 0.0001	Reject	0.734	1.010
	2	−1.356	0.011	161.8	< 0.0001	Reject	−1.565	−1.147
	3	−2.742	0.025	300.7	< 0.0001	Reject	−3.052	−2.432

where LL: lower limit and UL: upper limit.

Note that the three binomial models reject the null hypotheses H_0: $\beta_i = 0$, that is, their parameters are significant and fit the data well. Among them, the probit model is the one with the best fit for presenting higher values for the χ^2 statistic (which increases the power of the test). The predicted probabilities for each model are equal, regardless of the link function used, and their confidence intervals coincide in many cases, up to the order of thousandths. Now, suppose that after fitting any of these models, levels 2 and 3 of the A factor are added (see Table 3). Then, a new contingency table is obtained, such as the one shown in Table 6. In this case, the new model is given by

$$\begin{bmatrix}\eta_1^*\\ \eta_2^*\end{bmatrix} = \begin{bmatrix}g(p_1^*)\\ g(p_2^*)\end{bmatrix} = \begin{bmatrix}1\ 1\\ 1\ 0\end{bmatrix}\begin{bmatrix}\beta_1^*\\ \beta_2^*\end{bmatrix}.$$

The new estimates obtained by using the GUM, that is, re-adjusted binomial models using logit, probit and cloglog link functions on the resulting contingency table (see Table 6) are shown in Table 7. The predicted probabilities for each of the models, without attention to parameters significance levels, are reproduced in Table 8.

Given the last two levels aggregation, the new parameter vector $\boldsymbol{\beta}^*$ is estimated differently than in the original model, as shown in Table 7. As for the new predicted probability vector ($\widehat{\boldsymbol{p}}^*$), without considering the significance levels of the parameters, the results are as expected: the first component is the same as the original model ($\widehat{p}_1^* = \widehat{p}_1$), while in the second component you get that $\widehat{p}_2^* = 0.5257$, in accordance with Equation (16).

Table 5. Predicted probabilities and 95% CI according to link function.

	Estimation of p_i		95% CI	
Link	i	$\widehat{p}_i$	LL	UL
logit	1	0.4600	0.4084	0.5125
	2	0.1429	0.1100	0.1836
	3	0.9086	0.8736	0.9346
probit	1	0.4600	0.4083	0.5124
	2	0.1429	0.1093	0.1827
	3	0.9086	0.8746	0.9352
cloglog	1	0.4600	0.4094	0.5137
	2	0.1429	0.1102	0.1841
	3	0.9086	0.8755	0.9358

where LL: lower limit, and UL: upper limit.

Table 6. Example: $Y(0,1)$ versus $A(1,2)$.

	Y		
A	0	1	Total
1	189	161	350
2	332	368	700
Total	521	529	1050

Table 7. Usual procedure (GUM): $\widehat{\beta}_i^*$ and Wald test (H_0: $\beta_i^* = 0$) according to the link function.

		Estimation of β_i^*					95% CI		
Link	i	$\widehat{\beta}_i^*$	$V[\widehat{\beta}_i^*]$	χ^2	p-value	Decision	LL	UL	L(CI)
logit	1	0.103	0.006	1.8	0.174	Not reject	−0.045	0.251	0.296
	2	−0.263	0.017	4.0	0.045	Reject	−0.521	−0.006	0.515
probit	1	0.065	0.002	1.9	0.174	Not reject	−0.028	0.157	0.185
	2	−0.165	0.007	4.0	0.045	Reject	−0.326	−0.004	0.322
cloglog	1	−0.293	0.003	30.2	< 0.000	Reject	−0.398	−0.189	0.209
	2	−0.191	0.009	3.9	0.047	Reject	−0.380	−0.003	0.377

where L(CI): length of CI, LL: lower limit, and UL: upper limit.

Table 8. Usual procedure (GUM): Predicted probabilities and 95% CI according to link function

	Estimation of p_i^*		95% CI		
Link	i	$\widehat{p}_i^*$	LL	UL	L(CI)
logit	1	0.4600	0.4084	0.5125	0.1041
	2	0.5257	0.4887	0.5625	0.0738
probit	1	0.4600	0.4083	0.5124	0.1041
	2	0.5257	0.4887	0.5625	0.0738
cloglog	1	0.4600	0.4094	0.5137	0.1043
	2	0.5257	0.4893	0.5631	0, 0738

where L(CI): length of CI, Ll: lower limit, and UL: upper limit.

Although these are the expected values, with $\alpha = 0.05$ and according to the results of Table 7, in the logit and probit models there is insufficient evidence to reject H_0: $\beta_1^* = 0$, in a strict statistical sense, and so the predicted probabilities in Table 8 for these models should not be considered valid. In fact, the correct predictions for the logit model would be

$$\widehat{p}_1^* = \frac{\exp(\widehat{\beta}_1^* + \widehat{\beta}_2^*)}{1 + \exp(\widehat{\beta}_1^* + \widehat{\beta}_2^*)} = \frac{\exp(0 - 0.263)}{1 + \exp(0 - 0.263)} = \frac{\exp(-0.263)}{1 + \exp(-0.263)} = 0.4346 \quad \text{and}$$

$$\widehat{p}_2^* = \frac{\exp(\widehat{\beta}_1^*)}{1 + \exp(\widehat{\beta}_1^*)} = \frac{\exp(0)}{1 + \exp(0)} = 0.5000$$

while those corresponding to the probit model would be given by

$$\widehat{p}_1^* = \Phi(\widehat{\beta}_1^* + \widehat{\beta}_2^*) = \Phi(0 - 0.165) = \Phi(-0.165) = 0.4345 \quad \text{and}$$

$$\widehat{p}_2^* = \Phi(\widehat{\beta}_1^*) = \Phi(0) = 0.5000.$$

For that matter, only the cloglog model would remain valid since the nullity hypothesis for all its parameters is rejected.

In contrast, Table 9 and 10 present the estimates of the parameters and the predicted probabilities respectively, for the binomial models addressed, but now obtained through the GSM.

Table 9. Suggested procedure (GSM): $\widehat{\beta}_i^*$ and Wald test (H_0: $\beta_i^* = 0$) according to the link function.

		Estimation of β_i^*					95% CI		
Link	i	$\widehat{\beta}_i^*$	$V[\widehat{\beta}_i^*]$	χ^2	p-value	Decision	LL	UL	L(CI)
logit	1	0.103	0.002	4.5	0.034	Reject	0.008	0.198	0.190
	2	−0.263	0.014	5.0	0.025	Reject	−0.494	−0.033	0.462
probit	1	0.065	0.001	4.5	0.034	Reject	0.005	0.124	0.119
	2	−0.165	0.005	5.0	0.025	Reject	−0.309	−0.020	0.289
cloglog	1	−0.293	0.001	73.3	< 0.000	Reject	−0.360	−0.226	0.134
	2	−0.191	0.008	4.8	0.028	Reject	−0.362	−0.020	0.341

where L(CI): length of CI, Ll: lower limit, and UL: upper limit.

Table 10. Suggested procedure (GSM): Predicted probabilities and 95% CI according to link function.

	Estimation of p_i^*		95% CI		
Link	i	$\widehat{p}_i^*$	LL	UL	L(CI)
logit	1	0.4600	0.4084	0.5125	0.1041
	2	0.5257	0.5019	0.5494	0.0475
probit	1	0.4600	0.4083	0.5124	0.1041
	2	0.5257	0.5019	0.5494	0.0475
cloglog	1	0.4600	0.4094	0.5137	0.1043
	2	0.5257	0.5022	0.5497	0.0475

where L(CI): length of CI, Ll: lower limit, and UL: upper limit.

As expected by Equations (8) and (14), the estimates of β_i^* for each of the binomial models adjusted by using GUM and GSM are identical. However, the estimated variances are different in both procedures: those estimated by the GSM procedure are smaller. Therefore, it is preferable to those estimated by the GUM procedure. Of course, these decreases in the variances imply the reduction in the lengths of the confidence intervals, as it can be seen when comparing the respective tables. This applies both to the estimators of the linear predictor and to the second component of the predicted probability vector, regardless of the binomial model implemented.

Additionally, when the aggregate model data is adjusted by appealing to the suggested procedure, there is a change in the conclusions about the significance of $\widehat{\beta}_1^*$ for the logit and probit models. Thus, estimates of the predicted probabilities in Table 10 are now statistically valid and better approximate the available data. The latter also applies to the cloglog model which, even though their estimates are considered valid when obtained by the usual GUM procedure, improve when the GSM is used.

7. Conclusions

This work constitutes a generalization of the method proposed by Ponsot et al. (2009) for fitting binomial logit models, in the situation of factor levels aggregation on a simple contingency table (with a factor and a dichotomous response variable). This generalization consists of an extension, both of the usual procedure and of the procedure suggested by the author, for the adjustment of binomial models by means of link functions not only logit, but also probit and cloglog.

The results showed that the problem of factor levels aggregation persists in models that are usually competitors or alternatives to the logit model. That is because, regardless of the link function used, the violation of the binomial assumption remains when the associated probabilities of success at aggregated levels are dissimilar. Then, the suggested procedure maintains its advantages with any of the link functions used, being it preferable to the usual adjustment procedure, as it offers the necessary correction and subsequent improvement of the results.

The link function that is selected does not favor the application of a particular method between the two presented. However, it was confirmed that, in any scenario, that is, when any link function is appealed to derive a binomial model, the estimates obtained by the suggested method improve when the correct distribution assumption is used. The choice between one method or another is based mainly on the probabilities of success associated with the aggregate levels. If they present slight differences, the estimates of the usual method are not very different from those of the suggested method. On the contrary, as these differences grow, it is better to rely on the suggested method.

In the future, it is expected to apply the comparisons made for the saturated model, using different link functions, to the general case of the aggregation of factor levels in the unsaturated model, as proposed by Ponsot et al. (2012). It is also expected to apply re-sampling techniques to study the behavior of the standard errors of the estimators in this more general case. Extensions to the so-called tobit model are also of interest (Barros et al., 2018; Desousa et al., 2018). Finally, the applications of the methodology is explored in the sense proposed by Da Silva et al. (2016), that is, for multinomial-ordinal models of a longitudinal nature.

Acknowledgements

The authors would like to thank the Editors, Associate Editor and Referees for their suggestions and comments that led to a significant improvement of this manuscript.

Appendix A. Extreme values of $\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]$

Let $\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}] = n^*_{k-1}p^*_{k-1}(1-p^*_{k-1})$, where $n^*_{k-1} = n_{k-1} + n_k$ and $p^*_{k-1} = (n_{k-1}p_{k-1} + n_kp_k)/n^*_{k-1}$. Then, the partial derivatives of $\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]$, with respect to p_{k-1} and p_k, are given by

$$\begin{aligned}
\frac{\partial V_{\mathrm{Bin}}[Y^*_{k-1}]}{\partial p_{k-1}} &= n_{k-1}\left(1 - \frac{n_{k-1}p_{k-1} + n_kp_k}{n^*_{k-1}}\right) + (n_{k-1}p_{k-1} + n_kp_k)\left(-\frac{n_{k-1}}{n^*_{k-1}}\right) \\
&= \frac{n_{k-1}}{n^*_{k-1}}(n^*_{k-1} - 2n_{k-1}p_{k-1} - 2n_kp_k), \\
\frac{\partial V_{\mathrm{Bin}}[Y^*_{k-1}]}{\partial p_k} &= n_k\left(1 - \frac{n_{k-1}p_{k-1} + n_kp_k}{n^*_{k-1}}\right) + (n_{k-1}p_{k-1} + n_kp_k)\left(-\frac{n_k}{n^*_{k-1}}\right) \\
&= \frac{n_k}{n^*_{k-1}}(n^*_{k-1} - 2n_{k-1}p_{k-1} - 2n_kp_k).
\end{aligned}$$

Equating to zero, the critical points are all those ordered pairs of the form

$$(p_{k-1}, p_k) = \left(p_{k-1}, 0.5\left[1 + \frac{n_{k-1}}{n_k}(1-2p_{k-1})\right]\right).$$

The second partial derivatives of $\mathrm{V}_{\mathrm{Bin}}[Y^*_{k-1}]$ are stated as

$$\begin{aligned}
A &= \frac{\partial^2 V_{\mathrm{Bin}}[Y^*_{k-1}]}{\partial p^2_{k-1}} = \frac{n_{k-1}}{n^*_{k-1}}(-2n_{k-1}) = -\frac{2n^2_{k-1}}{n^*_{k-1}} \\
B &= \frac{\partial^2 V_{\mathrm{Bin}}[Y^*_{k-1}]}{\partial p_k \partial p_{k-1}} = \frac{n_{k-1}}{n^*_{k-1}}(-2n_k) = -\frac{2n_{k-1}n_k}{n^*_{k-1}} \\
C &= \frac{\partial^2 V_{\mathrm{Bin}}[Y^*_{k-1}]}{\partial p^2_k} = \frac{n_k}{n^*_{k-1}}(-2n_k) = -\frac{2n^2_k}{n^*_{k-1}}
\end{aligned}$$

while the Hessian is expressed by

$$H = AC - B^2 = \left(-\frac{2n^2_{k-1}}{n^*_{k-1}}\right)\left(-\frac{2n^2_k}{n^*_{k-1}}\right) - \left(-\frac{2n_{k-1}n_k}{n^*_{k-1}}\right)^2 = 0$$

Thus, that the Hessian criterion fails in the decision. Nevertheless, it is clear from the examination of Figure 1 that the set of points found are relative maximum values.

Appendix B. Extreme values of $\mathrm{V}[Y^*_{k-1}]$

Let $\mathrm{V}[Y^*_{k-1}] = n_{k-1}p_{k-1}(1-p_{k-1}) + n_kp_k(1-p_k)$. Then, the partial derivatives of $\mathrm{V}[Y^*_{k-1}]$, with respect to p_{k-1} and p_k, are given by

$$\begin{aligned}
\frac{\partial V[Y^*_{k-1}]}{\partial p_{k-1}} &= n_{k-1}(1-p_{k-1}) - n_{k-1}p_{k-1} = n_{k-1}(1-2p_{k-1}) \\
\frac{\partial V[Y^*_{k-1}]}{\partial p_k} &= n_k(1-p_k) - n_kp_k = n_k(1-2p_k)
\end{aligned}$$

which, when equated to zero, throw as a critical point $(p_{k-1}, p_k) = (0.5, 0.5)$.

As second partial derivatives of $\mathrm{V}[Y^*_{k-1}]$, we have

$$A = \frac{\partial^2 V[Y^*_{k-1}]}{\partial p^2_{k-1}} = n_{k-1}(-2) = -2n_{k-1}$$

$$B = \frac{\partial^2 V[Y^*_{k-1}]}{\partial p_k \partial p_{k-1}} = 0$$

$$C = \frac{\partial^2 V[Y^*_{k-1}]}{\partial p^2_k} = n_k(-2) = -2n_k.$$

Consequently, the Hessian is given by $H = AC - B^2 = (-2n_{k-1})(-2n_k) - (0)^2 = 4n_{k-1}n_k$. Due to $H > 0$ and $A < 0$, then at the critical point $(p_{k-1}, p_k) = (0.5, 0.5)$, there is a relative maximum whose value is $\mathrm{V}[Y^*_{k-1}] = (n_{k-1} + n_k)/4 = n^*_{k-1}/4$.

References

Agresti, A., 2007. An Introduction to Categorical Data Analysis. Wiley, NJ, US.

Agresti, A., 2015. Foundations of Linear and Generalized Linear Models. Wiley, New York, US.

Barros, M., Galea, M., Leiva, V., and Santos-Neto, M., 2018. Generalized tobit models: Diagnostics and application in econometrics. Journal of Applied Statistics, 45, 145–167.

Bonat, W.H., Lopes, J.E., Shimakura, S.E., and Ribeiro Jr, P.J., 2018. Likelihood analysis for a class of simplex mixed models. Chilean Journal of Statistics (ChJS), 9(2), 3–17.

Christensen, R., 1997. Log-Linear Models and Logistic Regression. Springer, NY, US.

Collet, D., 2002. Modelling Binary Data. Chapman and Hall, New York, US

Czado, C. and Munk, A., 2000. Noncanonical links in generalized linear models -when is the effort justified? Journal of Statistical Planning and Inference, 87, 317–345.

Czado, C. and Santner, T.J., 1992. The effect of link misspecification on binary regression inference. Journal of Statistical Planning and Inference, 33, 213–231.

Da Silva, N.B., Amorim, L.D., Colosimo, E.A., and Heller, L., 2016. Modeling ordinal longitudinal outcomes: an applied perspective of marginal and conditional approaches. Chilean Journal of Statistics, 7(2), 51–68.

Desousa, M., Saulo, H., Leiva, V., and Scalco, P., 2018. On a tobit-Birnbaum-Saunders model with an application to medical data. Journal of Applied Statistics, 45, 932–955.

Dobson, A., 2002. An Introduction to Generalized Linear Models. Chapman and Hall, New York, US

Hilbe, J.M., 2009. Logistic Regression Models. Chapman and Hall, New York, US

Hosmer, D., and Lemeshow, S., 2000. Applied Logistic Regression. Wiley, New York, US.

McCullagh, P., and Nelder, J., 1989. Generalized Linear Models. Chapman and Hall, New York, US.

McCulloch, C., and Searle, S., 2001. Generalized Linear, and Mixed Models. Wiley, New York, US.

Nelder, J.A., and Wedderburn, R.W.M., 1972. Generalized linear models. Journal of the Royal Statistical Society A, 135, 370–384.

Ponsot, E., 2011. Estudio de la agregación de niveles de los factores en el modelo logit binomial. Tesis de Doctorado, Universidad de Los Andes, Mérida, Venezuela.

Ponsot, E., Sinha, S., and Goitía, A., 2009. Sobre la agrupación de niveles del factor explicativo en el modelo logit binario. Revista Colombiana de Estadística, 32, 157–187.

Ponsot, E., Sinha, S., and Goitía, A., 2012. Aggregation of explanatory factor levels in a binomial logit model: Generalization to the multifactorial unsaturated case. Revista Colombiana de Estadística, 35, 139–166.

Tutz, G., 2011. Regression for Categorical Data. Cambridge University Press, Cambridge, UK.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: `chilean.journal.of.statistics@gmail.com`). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at `http://soche.cl/chjs/files/ChJS.zip`. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at `http://www.ams.org/mathscinet/msc/`. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright

Authors who publish their articles in the ChJS automatically transfer their copyright to the Chilean Statistical Society. This enables full copyright protection and wide dissemination of the articles and the journal in any format. The ChJS grants permission to use figures, tables and brief extracts from its collection of articles in scientific and educational works, in which case the source that provides these issues (Chilean Journal of Statistics) must be clearly acknowledged.