Revista Colombiana de Estadística
Junio 2012, volumen 35, no. 1, pp. 131 a 138

The Multinomial Logistic Model for the Case in which the Response Variable Can Assume One of Three Levels and Related Models

El modelo logístico multinomial para el caso en que la variable de respuesta puede asumir uno de tres niveles y modelos relacionados

Humberto Llinás[1,a], Carlos Carreño[2,b]

[1]Departamento de Matemática y Estadística, División de Ciencias básicas, Universidad del Norte, Barranquilla, Colombia

[2]Facultad de Ingeniería de Sistemas, Corporación Universitaria Americana, Barranquilla, Colombia

Abstract

The aim of this work is to examine multinomial logistic models when the response variable can assume three levels, generalizing a previous work of logistic models with binary response variables. We also describe some related models: The null, complete, and saturated models. For each model, we present and prove some theorems concerning to the estimation of the corresponding parameters with details that we could not find in the current literature.

***Key words*:** Binomial distribution, Logistic model, Multinomial logit.

Resumen

El objetivo de este trabajo es examinar los modelos de regresión logística multinomial cuando la variable de respuesta puede asumir tres niveles, generalizando un trabajo anterior con variables respuesta binarias. También describimos algunos modelos relacionados: los modelos nulo, completo y saturado. Para cada modelo, presentamos y demostramos teoremas relacionados con la estimación de los parámetros correspondientes con detalles que no fueron posibles encontrar en la literatura.

***Palabras clave*:** distribución binomial, logit multinomial, modelo logístico.

[a]Associate professor. E-mail: hllinas@uninorte.edu.co
[b]Lecturer. E-mail: ccarreno@coruniamericana.edu.co

1. Introduction

Llinás (2006) studied logistic models with a dichotomous response variable. A theorem was proved on the existence and uniqueness of maximum likelihood (ML) estimations for the logistic model and also about its calculations. Additionally, based on asymptotic theory for these ML-estimations and the score vector, approximations were found for different deviations $-2 \log L$, where L is the likelihood function. Based on these approximations, statistics were obtained for several hypothesis tests, each with an asymptotic chi-squared distribution. The asymptotic theory was developed for the case of independent, non-identically distributed variables; thus, modifications are required to apply this theory to the case of identically distributed variables. In this article, a distinction is always made between grouped data and ungrouped data.

Applications of the multinomial logistic model in various fields of engineering and health sciences have made this technique as a fundamental tool for data analysis and subsequent decision making. For this reason firstly, it is important to clarify the theoretical foundations of these models so that they can be applied to specific situations within the data analysis process, which requires more than the use of a statistical program.

We will present to the reader the theoretical background of this model in an effort to describe the continuity of its construction and the elements that are used to perform different analyses with respect to hypothesis tests, relative risks, odds, odds ratios, etc.

For this reason, and following the methodology proposed by Llinás (2006), this article studies multinomial logistic models only for the case in which the variable of interest can assume one of three levels. We describe related models, such as the null, full, and saturated models. For each model, the estimation theorems for the corresponding parameters are presented, providing details that are not found in the current literature (e.g., Agresti 1990, Hosmer & Lemeshow 2000, Kleinbaum & Klein 2002).

The article is organized into six sections. The first section consists of a introduction motivating this reason. The second section explains the basic Bernoulli model. The third section explains the full model. The fourth section explains the null model. The fifth section studies the saturated model and the basic assumptions, and the sixth section develops the theory corresponding to the multinomial logistic model.

2. The Bernoulli Model

Let us suppose that the variable of interest Y can assume one of three values or levels: 0, 1 or 2. For each $r = 0, 1, 2$, we let $p_r := P(Y = r)$ denote the probability that Y assumes the value r.

With n independent observations of Y, a sample $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ is obtained with data $y_i \in \{0, 1, 2\}$ for $i = 1, \dots, n$, in which y_i is a possible value of Y_i, which are independent of one another.

In order to construct the likelihood function, we create three independent binary variables with values of 0 and 1 as follows:

$$U_{ri} = \begin{cases} 1, & \text{if } Y_i = r \\ 0, & \text{otherwise} \end{cases}$$

where $r = 0, 1, 2$ and $i = 1, \dots, n$. Observe that $U_{ri} \sim \mathcal{B}(1, p_{ri})$, where $p_{ri} = P(Y_i = r)$.

In terms of the U_{ri} variables, the sample variables are $Y_i = (U_{0i}, U_{1i}, U_{2i})$, with values of $y_i = (u_{0i}, u_{1i}, u_{2i})$, in which $\sum\limits_{r=0}^{2} u_{ri} = 1$, for a fixed i. A statistical model is obtained in which

$$P(Y_i = y_i) = \prod_{r=0}^{2} p_{ri}^{u_{ri}}, \quad i = 1, \dots, n$$

Setting $\boldsymbol{y} = (y_1, \dots, y_n)^T$, we obtain the logarithm of the likelihood function for the $2n$-dimensional parameter $\boldsymbol{p} = (p_{01}, p_{11}, \dots, p_{0n}, p_{1n})^T$:

$$\mathcal{L}(\boldsymbol{p}) = \sum_{i=1}^{n} [u_{0i} \ln p_{0i} + u_{1i} \ln p_{1i} + (1 - u_{0i} - u_{1i}) \ln(1 - p_{0i} - p_{1i})] \tag{1}$$

3. The Complete Model

The *complete model* is characterized by the assumption that all p_{ri} (with $r = 0, 1, 2$ and $i = 1, \dots, n$) are considered parameters.

Theorem 1. *In the complete model, the ML-estimations of p_{ri} are $\widehat{P}_{ri} = U_{ri}$ with values $\widehat{p}_{ri} = u_{ri}$ for $r = 0, 1, 2$ and $i = 1, \dots, n$. Additionally, $\mathcal{L}_c := \mathcal{L}(y) = 0$.*

***Proof*.** Consider equation (1), in which

$$\mathcal{L}(\boldsymbol{p}) = \sum_{\substack{i \\ u_{0i}=1, u_{1i}=0}} \ln p_{0i} + \sum_{\substack{i \\ u_{0i}=0, u_{1i}=1}} \ln p_{1i} + \sum_{\substack{i \\ u_{0i}=0, u_{1i}=0}} \ln(1 - p_{0i} - p_{1i}).$$

Consider that $\mathcal{L}(\boldsymbol{p}) \overset{!}{=} 0$ if and only if $p_{0i} = u_{0i}$ and $p_{1i} = u_{1i}$ for each $i = 1, \dots, n$. This condition proves the existence of the ML-estimations. If for some i it is true that $p_{ri} \neq u_{ri}$, $r = 0, 1$, then $\mathcal{L}(\boldsymbol{p}) < 0$. This condition demonstrates that the ML-estimations are unique because if $\widetilde{\boldsymbol{p}}$ is a vector that has at least one p_{ri} component that is different from u_{ri}, then $\mathcal{L}(\boldsymbol{p}) < \mathcal{L}_c$ (given that upon replacing $p_{ri} = u_{ri}$ in $\mathcal{L}(\boldsymbol{p})$, $\mathcal{L}_c = 0$). □

4. The Null Model

The *null model* is characterized by the assumption that for each $r = 0, 1, 2$, all the p_{ri} values ($i = 1, \ldots, n$) are considered equal; that is, there are two parameters, p_0 and p_1. In this case, equation (1) becomes

$$\mathcal{L}(\boldsymbol{p}) \quad = \quad n[\overline{u}_0 \ln p_0 + \overline{u}_1 \ln p_1 + (1 - \overline{u}_0 - \overline{u}_1) \ln(1 - p_0 - p_1)] \tag{2}$$

in which $\overline{u}_r = \sum_{i=1}^{n} \frac{u_{ri}}{n}$.

Theorem 2. *In the null model, the ML-estimation of* p_r *is* $\widehat{P}_r = \overline{U}_r$ *with value* $\widehat{p}_r = \overline{u}_r$*. Additionally,* $\mathcal{L}_o := \mathcal{L}(\widehat{p}) < 0$ *if and only if* $0 < \overline{u}_0 + \overline{u}_1 < 1$.

***Proof*.** It is clear that from equation (2) that

- If $\overline{u}_0 + \overline{u}_1 = 0$, then $\overline{u}_0 = \overline{u}_1 = 0$. Therefore, $\mathcal{L}(p) = 0$ if and only if $p_r = 0 = \overline{u}_r$.
- If $\overline{u}_0 + \overline{u}_1 = 1$, then $\overline{u}_0 = 0$ or $\overline{u}_1 = 0$. Therefore, for $\overline{u}_0 = 0$, $\mathcal{L}(p) = 0$ if and only if $p_1 = 1 = \overline{u}_1$ and $\overline{u}_1 = 1$, $\mathcal{L}(p) = 0$ if and only if $p_0 = 1 = \overline{u}_0$.
- Now let us assume that $0 < \overline{u}_0 + \overline{u}_1 < 1$. From equation (2) and for a given r, it can be proven that

  $$\frac{\partial \mathcal{L}(\boldsymbol{p})}{\partial p_r} \; = \; \frac{\overline{u}_r}{p_r} - \frac{1 - \overline{u}_0 - \overline{u}_1}{1 - p_0 - p_1} \; = \; 0$$

  if and only if $\widehat{p}_r = \overline{u}_r$. Given that

  $$\frac{\partial^2 \mathcal{L}(\widehat{\boldsymbol{p}})}{\partial p_r^2} \; = \; -\left[\frac{\overline{u}_r}{\widehat{p}_r^2} - \frac{1 - \overline{u}_0 - \overline{u}_1}{(1 - \widehat{p}_0 - \widehat{p}_1)^2}\right] \; < \; 0$$

  this solution is unique. Additionally, $\ln \overline{u}_r$ and $\ln(1 - \overline{u}_0 - \overline{u}_1)$ are both negative. Therefore, $\mathcal{L}_o < 0$. □

5. The Saturated Model and Assumptions

The saturated model is characterized by the following assumptions:

Assumption 1. It is assumed that:

1. There are K explanatory variables $X_1, \ldots, X_K$ (some may be numerical and other may be categorical) with values $x_{1i}, \ldots, x_{Ki}$ for $i = 1, \ldots, n$ (which are set or observed by the statistician depending on whether the variables are deterministic or random);
2. Among the n individual vectors $(x_{1i}, \ldots, x_{Ki})$ of the values of the explanatory variables Xs, there are J different possible combinations, defining J populations. Therefore, $J \leq n$. J is often referred to as the number of covariate patterns in the applied literature.

Notation. The notation for each population $j = 1, \dots, J$ is denoted as follows:

- The number of Y_{ij} observations (or of U_{rij} observations in the r category) in each jth population is n_j, with $n_1 + \cdots + n_J = n$;
- For a fixed $r = 0, 1, 2$; the random variable corresponding to the sum of the n_j observations of U_{rij}, given by $Z_{rj} := \sum_{i=1}^{n_j} U_{rij}$ with value $z_{rj} = \sum_{i=1}^{n_j} u_{rij}$, in which $\sum_{j=1}^{J} z_{rj} = \sum_{i=1}^{n} u_{ri}$.

For simplicity, the jth population $(x_{1j}, \dots, x_{Kj})$ will be abbreviated with the symbol $\star$.

Assumption 2. For each fixed $r = 0, 1, 2$, each population $j = 1, \dots, J$ and each observation $i = 1, \dots, n$ in population j, it is assumed that

- $(U_{rij} \mid \star) \sim \mathcal{B}(1, p_{rj})$
- The $(U_{rij} \mid \star)$ variables are independent of one another.

Below, the $\star$ symbol will be omitted. Assumption 2 implies the following:

1. For each $r = 0, 1, 2$ and each fixed $j = 1, \dots, J$, all the p_{rij}, $i = 1, \dots, n$, in each jth population are equal. In other words, the $2J$-dimensional $p = (p_{01}, p_{11}, \dots, p_{0J}, p_{1J})^T$ vector is the parameter.
2. For each $r = 0, 1, 2$ and each population $j = 1, \dots, J$:
   - $Z_{rj} \sim \mathcal{B}(n_j, p_{rj})$
   - The Z_{rj} variables are independent among populations.

In the saturated model, the logarithm of the maximum likelihood function will be

$$\mathcal{L}(\boldsymbol{p}) = \sum_{j=1}^{J} [z_{0j} \ln p_{0j} + z_{1j} \ln p_{1j} + (n_j - z_{0j} - z_{1j}) \ln(1 - p_{0j} - p_{1j})] \tag{3}$$

Theorem 3. *In the saturated model, the ML-estimations of* p_{rj} *are* $\widetilde{P}_{rj} = \frac{Z_{rj}}{n_j}$, *with the values* $\widetilde{p}_{rj} = \frac{z_{rj}}{n_j}$, $j = 1, \dots, J$. *Furthermore,*

$$\mathcal{L}(\widetilde{\boldsymbol{p}}) = \sum_{j=1}^{J} n_j [\widetilde{p}_{0j} \ln \widetilde{p}_{0j} + \widetilde{p}_{1j} \ln \widetilde{p}_{1j} + (1 - \widetilde{p}_{0j} - \widetilde{p}_{1j}) \ln(1 - \widetilde{p}_{0j} - \widetilde{p}_{1j})] \tag{4}$$

It also holds that $\mathcal{L}_s := \mathcal{L}(\widetilde{\boldsymbol{p}}) < 0$ *for* $0 < \widetilde{p}_{rj} < 1$.

***Proof*.** Let us hold r and j. If $0 < \widetilde{p}_{rj} < 1$, then we have

$$\frac{\partial \mathcal{L}}{\partial p_{rj}} = \frac{z_{rj}}{p_{rj}} - \frac{n_j - z_{0j} - z_{1j}}{1 - p_{0j} - p_{1j}} = 0$$

if and only if $\widetilde{p}_{rj} = \frac{z_{rj}}{n_j}$. Therefore, if $0 < z_{rj} < n_j$, for each r and j, then we have

$$\left.\frac{\partial^2 \mathcal{L}}{\partial p_{rj}^2}\right|_{p_{rj}=\widetilde{p}_{rj}} = -\left[\frac{n_j^2}{z_{rj}} + \frac{n_j^2}{n_j - z_{0j} - z_{1j}}\right] < 0$$

Two extreme cases must be analyzed:

- If $z_{rj} = 0$, then $\frac{\partial \mathcal{L}}{\partial p_{rj}} = -\frac{n_j}{1-p_{0j}-p_{1j}}$ decreases in p_j. In this case, $\mathcal{L}$ decreases in p_{rj}; that is, $\mathcal{L}(\boldsymbol{p})$ is maximized for $p_{rj} = 0$.
- If $z_{rj} = n_j$, then, $\frac{\partial \mathcal{L}}{\partial p_j} = \frac{n_j}{p_j}$ increases in p_{rj}. In this case, $\mathcal{L}$ increases in p_{rj}; that is, $\mathcal{L}(\boldsymbol{p})$ is maximized for $p_{rj} = 1$.

In the saturated model, the value of $\mathcal{L}$ can be obtained by replacing in equation (3), each p_{rj} with $\widetilde{p}_{rj}$, $j = 1, \ldots, J$. Thus, we obtain equation (4). Under the condition that $0 < \widetilde{p}_{rj} < 1$ it can be shown that $\ln \widetilde{p}_{rj}$ y $\ln(1 - \widetilde{p}_{rj})$ are both negative. Therefore, the sum of the right side of equation (4) is also negative. □

6. The Multinomial Logistic Model

6.1. Assumptions

Assumptions 1 and 2 from section 5 are preserved, with the additional assumption that a matrix

$$\boldsymbol{C} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{K1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1J} & \cdots & x_{KJ} \end{pmatrix}$$

has a complete range $Rg(C) = 1 + K \leq J$. To obtain a logistic model, one of the categories of the dependent variable Y, such as 0, is used as a reference. The following additional assumption is also made:

Assumption 3.

$$g_1(\boldsymbol{x}_j) = \ln\left(\frac{p_{1j}}{p_{0j}}\right) = \delta_1 + \beta_{11}x_{j1} + \cdots + \beta_{1K}x_{jK} \tag{5}$$

$$g_2(\boldsymbol{x}_j) = \ln\left(\frac{p_{2j}}{p_{0j}}\right) = \delta_2 + \beta_{21}x_{j1} + \cdots + \beta_{2K}x_{jK} \tag{6}$$

in which $\boldsymbol{x}_j := (1, x_{j1}, \ldots, x_{jK})^T$. Let

$$\boldsymbol{\alpha} = (\delta_1, \beta_{11}, \ldots, \beta_{1K}, \delta_2, \beta_{21} \ldots, \beta_{2K})^T$$

denote the vector of the $2(1+K)$ parameters in the model. Note that the assumption that $Rg(\boldsymbol{C}) = 1 + K$ allows the $\boldsymbol{\alpha}$ parameter to be identified.

For a given observation x_j in population j and for the so-called risk is calculated as follows:

$$p_{rj} = \frac{\exp\{g_r(x_j)\}}{\sum_{s=0}^{2} \exp\{g_s(x_j)\}} \tag{7}$$

for each $r = 0, 1, 2$ and with $g_0(x_j) = 0$. The logarithm of the likelihood function can be written as a function of α, as follows:

$$\mathcal{L}(\boldsymbol{\alpha}) = \sum_{j=1}^{J} \left[z_1 g_1(x_j) + (n_j - z_{0j} - z_{1j}) g_2(x_j) - n_j \ln \left(\sum_{r=0}^{2} \exp\{g_r(x_j)\} \right) \right] \tag{8}$$

6.2. Relation between the Multinomial Logistic Model and the Saturated Model

The equations of assumption 3 in Section 6.1 can be written in a vector form, where $\boldsymbol{g}_r = \boldsymbol{C}\boldsymbol{\beta}_r$, $r = 1, 2$, in which $\boldsymbol{g}_r$ is a J-dimensional vector with elements $g(x_j)$, $j = 1, 2, \ldots, J$.

Given the above, the following cases can be highlighted:

Case 1. $J = 1 + K$

In this case, $\boldsymbol{C}$ is an invertible matrix. Therefore,

$$\beta_r = \boldsymbol{C}^{-1} g_r, \quad r = 1, 2$$

That is, there is a one-to-one relationship between the parameters of the saturated model and those of the logistic model. In other words, the two models express the same thing.

Particularly, the ML-estimations of the probabilities p_{rj} are equal in both models: $\widehat{p}_{rj} = \widetilde{p}_{rj}$ for each $j = 1, 2, \ldots, 1 + K$.

Case 2. $J > 1 + K$

In this case, $\widehat{\alpha}$ must first be calculated, and based on these values, the p_{rj} values can be calculated. In general, we observe that $\widehat{p}_{rj} \neq \widetilde{p}_{rj}$.

7. Likelihood Equations

The likelihood equations are found by calculating the first derivatives of $\mathcal{L}(\boldsymbol{\alpha})$ with respect to each one of the $2(1 + K)$ unknown parameters, as follows. For every $k = 0, 1, \ldots, K$, we have

$$\frac{\partial \mathcal{L}(\boldsymbol{\alpha})}{\partial \beta_{1k}} = \sum_{j=1}^{J} \left[z_{1j} x_{jk} - \frac{n_j \, x_{jk} \, e^{\boldsymbol{g}_1(x_j)}}{1 + e^{\boldsymbol{g}_1(x_j)} + e^{\boldsymbol{g}_2(x_j)}} \right] = \sum_{j=1}^{J} x_{jk} (z_{1j} - n_j p_{1j})$$

and

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{\alpha})}{\partial \beta_{2k}} &= \sum_{j=1}^{J}\left[(n_j - z_{0j} - nz_{1j})x_{jk} - \frac{n_j\, x_{jk}\, e^{\boldsymbol{g}_2(x_j)}}{1 + e^{\boldsymbol{g}_1(x_j)} + e^{\boldsymbol{g}_2(x_j)}}\right] \\ &= \sum_{j=1}^{J} x_{jk}[(n_j - z_{0j} - z_{1j}) - n_j p_{2j}] \\ &= \sum_{j=1}^{J} x_{jk}(z_{2j} - n_j p_{2j})\end{aligned}$$

Therefore, for every $k = 0, 1, \ldots, K$ and every $r = 0, 1, 2$, the likelihood equations are given by

$$\frac{\partial \mathcal{L}(\boldsymbol{\alpha})}{\partial \beta_{rk}} = \sum_{j=1}^{J} x_{jk}(z_{rj} - n_j p_{rj})$$

The estimator of maximum likelihood is obtained by setting these equations equal to zero and solving for the logistic parameters. The solution requires the same type of iterations that were used to obtain the estimations in the binary case, as demonstrated in Llinás (2006).

8. Conclusions

We have studied the multinomial logistic models when the response variable can assume one of three values and also described some related models such as the null, complete, and saturated models. We have presented and proved the theorems 1, 2 and 3, which give us the estimation of the corresponding parameters.

[Recibido: junio de 2011 — Aceptado: febrero de 2012]

References

Agresti, A. (1990), *Categorical Data Analysis*, 2 edn, John Wiley and Sons, Inc., New York.

Hosmer, D. & Lemeshow, S. (2000), *Applied Logistic Regresssion*, 2 edn, John Wiley and Sons, Inc., New York.

Kleinbaum, D. & Klein, M. (2002), *Logistic Regression: A Self-Learning Text*, 2 edn, Springer, New York.

Llinás, H. (2006), 'Precisiones en la teoría de los modelos logísticos', *Revista Colombiana de Estadística* **29**(2), 239–265.