1. Classificação *INPE-COM.9/PRE*
CDU: 330.115

2. Período
Julho/81

4. Distribuição

interna ☐

externa ☒

3. Palavras Chaves (selecionadas pelo autor)
ESTIMAÇÃO
MODELO LOGIT
VARIÁVEIS QUALITATIVAS

5. Relatório nº
INPE-2218-PRE/016

6. Data
Setembro, 1981

7. Revisado por
E. A. N. Lorena

8. Título e Sub-Título
ESTIMAÇÃO EM MODELOS ENVOLVENDO VARIÁVEIS DEPENDENTES QUALITATIVAS

9. Autorizado por
Nelson de J. Parada
Diretor

10. Setor *DAT/DSE* Código

11. Nº de cópias *10*

12. Autoria *Nívea Teixeira Dias*
Múcio Roberto Dias

14. Nº de páginas *20*

15. Preço

13. Assinatura Responsável

16. Sumário/Notas

Neste trabalho são apresentados vários métodos para estimar os parâmetros do modelo logit (ou logístico), empregado na análise do comportamento de variáveis aleatórias discretas. Os métodos apresentados são comparados e as vantagens e desvantagens de cada um, bem como as condições em que eles são aplicáveis, são discutidas.

17. Observações
Este trabalho foi apresentado na 33a. Reunião Anual da SBPC, realizada em Salvador de 8 a 15 de julho de 1981.

ÍNDICE

ABSTRACT

In this paper several methods for estimating the logit (or logistic) model used to analyse the behaviour of discrete random variables are presented. The advantages and disadvantages of each method, as well as the aplicability of each one, are discussed.

CAPÍTULO I

INTRODUÇÃO

Um dos problemas fundamentais relativos ao estudo de modelos de decisão é entender o comportamento humano frente à escolha. Formulam-se modelos ou hipóteses sobre a natureza do processo de decisão e, então, avaliam-se essas hipóteses à luz do comportamento observado. Nesse estudo, uma dificuldade encontrada é que muitos dos fenômenos comportamentais de interesse são de natureza qualitativa, não sendo os métodos estatísticos comumente usados no estudo de relações entre variáveis apropriados para estimar tais tipos de relação.

A análise estatística de modelos com variáveis dependentes qualitativas pode ser vista como o problema de predizer as probabilidades associadas aos vários valores da variável dependente. Essa análise requer o uso de técnicas que ainda não estão bem desenvolvidas; no entanto, ultimamente a área tem recebido uma atenção significativa, como evidenciado por Nerlove e Press (1973).

Pode-se dizer que a análise de dados qualitativos é constituída de duas fases; a primeira compreende a especificação de uma forma funcional para as probabilidades básicas; a segunda abrange a consideração da técnica apropriada de estimação do modelo especificado.

Neste trabalho, faz-se a apresentação detalhada do Modelo Logit, caracterizado pela especificação das probabilidades básicas em termos da chamada função logística, definida como:

$$F(t) = \frac{e^t}{1 + e^t} \quad , \qquad t \in (-\infty, +\infty) \tag{I.1}$$

Apresentam-se, também, alguns métodos de estimação comumente usados na análise de variáveis qualitativas, salientando-se as condições de aplicabilidade de cada um deles.

Especificamente, os métodos abordados são: mínimos quadrados ordinários, mínimos quadrados ponderados , qui-quadrado logit mínimo e máxima verossimilhança.

CAPÍTULO II

O MODELO LOGIT

Seja Y uma variável aleatória definida como:

$$Y = \begin{cases} 1 & \text{se um determinado evento A ocorre} \\ 0 & \text{caso contrário} \end{cases}$$

e seja P a probabilidade de uma resposta positiva, ou seja, $P \triangleq P(Y=1) = P(A)$.

O modelo Logit supõe que a probabilidade de uma resposta positiva para o j-ésimo indivíduo está relacionada a um vetor X_j de variáveis não-estocásticas, através da função F(.) definida na Equação I.1, ou seja:

$$P_j = P(Y_j = 1 | X_j) = F(X_j\beta) = \frac{e^{X_j\beta}}{1+e^{X_j\beta}} \qquad \text{ou}$$

$$P_j = \frac{1}{1 + e^{-X_j\beta}} \;,\; j = 1,2,\ldots., N \qquad \text{II.1}$$

onde β é o vetor de parâmetros a serem estimados, e N é o tamanho da amostra. A função logística é muito usada, sendo sua popularidade decorrente de suas convenientes propriedades matemáticas.

Note-se que, se $P_j = \dfrac{1}{1 + e^{-X_j\beta}}$, então

$$1 - P_j = \frac{e^{-X_j\beta}}{1 + e^{-X_j\beta}} \;,$$ de tal forma que

$$\frac{P_j}{1 - P_j} = e^{X_j\beta} \qquad \text{II.2}$$

Tomando-se logaritmos em ambos os lados da Equação 2, tem-se:

$$\ell g \left[\frac{P_j}{1 - P_j}\right] = X_j\beta \quad , \; j = 1,2,\ldots, N \qquad \text{II.3}$$

A expressão $\ell g \left[\frac{P_j}{1 - P_j}\right]$, denotada por L_j, é conhecida como a *"logit"* da probabilidade P_j; este termo foi criado por Berkson (1953) e designa o logaritmo natural dos *"odds"* de uma resposta positiva para o j-ésimo indívíduo.

Vê-se que, enquanto $X_j\beta$ varia de $-\infty$ a $+\infty$, P_j varia entre 0 e 1 e L_j varia de $-\infty$ a $+\infty$. Assim, enquanto as probabilidades são limitadas, as logits são ilimitadas com relação aos valores de X_j. Além disso, L_j é uma função monotonicamente crescente de P_j; pode, portanto, ser usada como uma medida de incerteza, do mesmo modo que a probabilidade P_j.

Considere-se, agora, o caso em que um indivíduo escolhe entre três ou mais alternativas. Por exemplo, ao escolher uma maneira de se dirigir ao trabalho, o indivíduo pode optar por condução própria, ônibus, metrô, etc. Em cada caso, supõe-se que as alternativas são mutuamente excludentes, isto é, somente uma escolha é feita. Além disso, considera-se, também, que não existe uma ordem natural associada ao conjunto de valores que denotam as possíveis alternativas.

Nesse caso, as funções estimadas expressam o logaritmo natural dos *"odds"* de um resultado em relação a um outro qualquer, como uma função linear das variáveis explicativas. Este logaritmo é denominado logit condicional.

Sejam Y_1, $Y_2,\dots,Y_N$ observações independentes de uma variável discreta Y, que pode assumir um entre m valores diferentes a_1, a_2, ..., a_m. A cada par (k,ℓ), $k,\ell = 1,2,\dots,m$, $k \neq \ell$, é associado um vetor de parâmetros $\beta_{k\ell}$ que deve ser estimado.

Definindo-se P_{jk} como

$$P_{jk} \triangleq P(Y_j = a_k), \quad \begin{array}{l} k = 1,\dots, m \\ j = 1,2,\dots,N \end{array}$$

tem-se:

$$\sum_{k=1}^{m} P_{jk} = 1 \quad , \quad j = 1,2,\dots,N \qquad \text{II.4}$$

O modelo Logit é, então, escrito como:

$$\ell g \left(\frac{P_{jk}}{P_{j\ell}} \right) = X_j\beta_{k\ell} = L_{jk} \ , \quad \begin{array}{l} j = 1,2,\dots,N \\ k,\ell = 1,2,\dots,m \\ k \neq \ell \end{array} \qquad \text{II.5}$$

O modelo acima expressa as logits condicionais, denotadas por L_{jk}, como funções lineares de X_j, onde X_j é o vetor de variáveis explicativas associadas ao j-ésimo indivíduo.

Pode-se observar que os parâmetros no lado direito da Equação 5 estão sujeitos a restrições. Considere-se, por exemplo, essa equação para os *"odds"* $\frac{P_{j\ell}}{P_{ji}}$:

$$\ell g \left(\frac{P_{j\ell}}{P_{ji}} \right) = X_j\beta_{\ell i} \quad , \quad j = 1,\dots,N$$

Se a equação acima é somada à Equação 5, tem-se:

$$\ell g\left(\frac{P_{jk}}{P_{j\ell}}\right) + \ell g\left(\frac{P_{j\ell}}{P_{ji}}\right) = X_j\beta_{k\ell} + X_j\beta_{\ell i} = X_j\ (\beta_{k\ell} + \beta_{\ell i})$$

mas,

$$\ell g\left(\frac{P_{jk}}{P_{j\ell}}\right) + \ell g\left(\frac{P_{j\ell}}{P_{ji}}\right) = \ell g\left(\frac{P_{jk}}{P_{ji}}\right) = X_j\beta_{ki}$$

Portanto, tem-se:

$$\beta_{ki} = \beta_{k\ell} + \beta_{\ell i} \quad => \quad \beta_{k\ell} = \beta_{ki} - \beta_{\ell i}$$

Como salienta Theil (1970) estas condições de *"circularidade"* implicam que $\beta_{k\ell}$ pode ser escrito na forma $\beta_k - \beta_\ell$, para valores apropriados desses β's. Pode-se, portanto, escrever a Equação 5 na forma:

$$\ell g\left(\frac{P_{jk}}{P_{j\ell}}\right) = X_j\ (\beta_k - \beta_\ell) \qquad \text{II.6}$$

$j = 1,2,\ldots,N$
$k,\ell = 1,2,\ldots,m$
$k \neq \ell$

Vê-se, então, que é suficiente considerar a Equação 6 para um único valor de ℓ, por exemplo, $\ell = 1$, e para todos os valores de k diferentes de 1; as equações para todas as outras combinações (k, ℓ) podem ser obtidas a partir destas.

Fazendo-se $\beta_1 = \beta_{11} = 0$, o modelo se resume a:

$$L_{jk} = \ell g \left(\frac{P_{jk}}{P_{j1}} \right) = X_j \beta_k \qquad \begin{array}{l} j = 1,2,\ldots,N \\ k = 2,3,\ldots,m \end{array} \qquad \text{II.7}$$

As m-1 equações na Expressão 7 mais a restrição da Equação 4 determinam as probabilidades univocamente. Estas são:

$$P_{j1} = \frac{1}{1 + \sum_{\ell=2}^{m} e^{X_j \beta_\ell}}$$

$$P_{jk} = \frac{e^{X_j \beta_k}}{1 + \sum_{\ell=2}^{m} e^{X_j \beta_\ell}} , \qquad \begin{array}{l} j = 1,2,\ldots,N \\ k = 2,3,\ldots,m \end{array}$$

ou, genericamente:

$$P_{jk} = \frac{e^{X_j \beta_k}}{\sum_{\ell=1}^{m} e^{X_j \beta_\ell}} \qquad \begin{array}{l} j = 1,2,\ldots,N \\ k = 1,2,\ldots,m \end{array} \qquad \text{II.8}$$

CAPÍTULO III

ESTIMAÇÃO DO MODELO

Para estimar os parâmetros da equação $L_{jk} = X_j\beta_k$ existem vários métodos; dentre eles, tratar-se-ão, neste trabalho, dos seguintes:

- método de mínimos quadrados ordinários;
- método de mínimos quadrados ponderados;
- método do qui-quadrado logit mínimo; e
- método de máxima verossimilhança.

Para descrever estes métodos, consederar-se-á, sem perda de generalidade, o caso de uma variável binária (m=2). Para o caso mais geral, m>2, vejam-se Nerlove e Press (1973).

3.1 - O MÉTODO DE MÍNIMOS QUADRADOS ORDINÁRIOS

Dispondo-se de várias observações de Y para cada valor de X, é possível, dado X, calcular as frequências relativas de cada resultado e usá-las como estimativas da probabilidade condicional de ocorrência do evento de interesse.

Considere-se, por exemplo, uma amostra que consiste em n_1 famílias com renda X_1, n_2 famílias com renda X_2, etc. Seja N o número de *"níveis"* de renda e defina-se a variável Y como:

$$Y_{ij} = \begin{cases} 1 & \text{se a i-ésima família no grupo j mora em casa} \\ & \text{própria (evento A ocorre).} \\ 0 & \text{caso contrário.} \end{cases}$$

$i = 1,2,\ldots,n_j$
$j = 1,2,\ldots,N$

Seja f_j a frequência relativa de ocorrência do evento A no grupo j, ou seja:

$$f_j = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} \quad , \quad j = 1,2,\ldots,N \qquad \text{III.1}$$

Definindo-se a logit observada ℓ_j como

$$\ell_j = \ell g \left(\frac{f_j}{1 - f_j} \right) \qquad \text{III.2}$$

e adotando-se o modelo:

$$\ell_j = X_j\beta + \varepsilon_j \quad , \qquad \text{III.3}$$

onde $E(\varepsilon_j) = 0$, $V(\varepsilon_j) = \alpha^2$, $E(\varepsilon_j\varepsilon_k) = 0$, $j \neq k$, e X_j é o vetor de variáveis explicativas para o grupo j, obtém-se o estimador de mínimos quadrados de β, dado por:

$$\hat{\beta} = (X'X)^{-1} \; X' \; \ell$$

onde X é a matriz de valores observados das variáveis independentes e ℓ é o vetor das logits observadas.

A aproximação de P_j por f_j é razoável uma vez que, como se sabe, $\sum_{i=1}^{n_j} Y_{ij}$ tem distribuição binomial, para a qual a frequência média de ocorrência do evento A é f_j.

O método de mínimos quadrados ordinários fornece estimativas consistentes somente quando o número de repetições para cada um dos níveis de x for suficientemente grande, pois, nesse caso, a variável dependente tem distribuição aproximadamente normal.

Uma dificuldade surge com o método quando há apenas uma observação de Y para cada nível de x, ou seja, quando n_j=1. Neste caso, f_j = 0 ou 1, os *"odds"* são 0 ou ∞ e as logits não são sequer definidas. Existem algumas regras para usar o método descrito, por exemplo, que n_j seja no mínimo 5 (Hanushek e Jackson (1977)). No entanto, uma regra precisa deve levar em conta o fato de que a aproximação é mais pobre para níveis de x em que a frequência de uma dada escolha é próxima de 0 ou de 1.

Além das dificuldades mencionadas acima, surgem também problemas quando as variáveis explicativas são contínuas, visto que, nesse caso, espera-se a observação de um só valor de Y para cada valor de X. Este problema é contornado agrupando-se os valores de X em categorias, de modo que se obtenha um número maior de valores de Y em cada nível.

3.2 - O MÉTODO DE MÍNIMOS QUADRADOS PONDERADOS

Considere-se o modelo definido na Equação 3.

Sob as suposições binomiais usuais, a frequência relativa f_j é assintoticamente normal com média P_j e variância $\frac{P_j\,(1 - P_j)}{n_j}$. Como ℓ_j é uma função de f_j que satisfaz a certas condições de regularidade, ℓ_j é assintoticamente normal com média L_j e variância assintotica aproximadamente igual a $\frac{1}{n_j\,P_j\,(1-P_j)}$.

O método de mínimos quadrados ponderados propõe o uso das recíprocas dessas variâncias como pesos. Assim, dados os valores f_j, peso mais alto é dado aos níveis de X_j nos quais o número de observações (n_j) é maior. Além disso, níveis com frequência relativa igual a 0 ou 1 têm peso 0, o que é importante porque, nesses casos, a logit observada ℓ_j não é definida. Similarmente, dado n_j, o peso é pequeno quando f_j é próxima de 0 ou 1, o que é razoável, já que ℓ_j assume, nessas condições, valores muito grandes (positivos ou negativos), sendo altamente sensível a pequenas mudanças em f_j. Assim, parece conveniente atribuir peso baixo a tais observações.

Assim, o método em questão busca minimizar a expressão:

$$(\ell - X\beta)' \Omega^{-1} (\ell - X\beta)$$

onde ℓ e X são definidos como na seção anterior, e Ω é a matriz de covariância dos ε's, dada por:

$$\Omega = \begin{pmatrix} \frac{1}{n_1 P_1(1 - P_1)} & 0 & 0 \\ 0 & \frac{1}{n_2 P_2(1 - P_2)} & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \frac{1}{n_N P_N(1 - P_N)} \end{pmatrix}$$

Para aplicar esse procedimento, são necessários os P_j's. Como estes são desconhecidos, usam-se as frequências f_j como estimativas dos primeiros, obtendo-se, assim, a matriz de covariância estimada $\hat{\Omega}$, que é:

$$\hat{\Omega} = \begin{pmatrix} \frac{1}{n_1 f_1(1 - f_1)} & 0 & 0 \\ 0 & \frac{1}{n_2 f_2(1 - f_2)} & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \frac{1}{n_N f_N(1 - f_N)} \end{pmatrix}$$

Para n_j suficientemente grande ($j = 1,2,\ldots,N$), pode-se substituir $\underset{\sim}{\omega}$ por $\hat{\underset{\sim}{\omega}}$ uma vez que, neste caso, f_j é um estimador consistente de P_j.

Obtém-se, dessa maneira, uma estimativa de β, computada através de uma regressão ponderada com pesos $\omega_j = n_j\, f_j(1 - f_j)$. A estimativa obtida é:

$$\hat{\beta} = (X'\hat{\underset{\sim}{\omega}}^{-1} X)^{-1} X'\hat{\underset{\sim}{\omega}}^{-1} \ell$$

A deficiência principal desse método está em que níveis com f_j igual a 0 ou 1, bem como aqueles níveis que contêm somente uma ou nenhuma observação, são ignorados. Se $n_j > 1$, isto pode constituir uma característica indesejável do método. Além disso, o método se baseia em resultados assintóticos e é, portanto, aplicável somente a grandes amostras.

3.3 - O MÉTODO QUI-QUADRADO LOGIT MÍNIMO

Este método foi desenvolvido por Berkson (1953) e sua idéia básica consiste em minimizar a estatística X^2(logit), definida como:

$$X^2(\text{logit}) = \sum_{j=1}^{N} \frac{[f_j - E(f_j)]^2}{V(f_j)}$$

onde $E(f_j) = P_j$, $V(f_j) = \dfrac{P_j(1 - P_j)}{n_j}$

Como $P_j = \dfrac{1}{1 + e^{-X_j\beta}}$, tem-se

$$X^2(\text{logit}) = \sum_{j=1}^{N} \frac{n_j(f_j - P_j)^2}{P_j(1 - P_j)} = \sum_{j=1}^{N} n_j(f_j + f_j e^{-X_j\beta} - 1)^2 e^{X_j\beta}$$

O procedimento acima possui algumas propriedades estatísticas interessantes. Em primeiro lugar, a estatística X^2(logit) tem distribuição assintótica de qui-quadrado. Em segundo lugar, as estatísticas fornecidas são assintoticamente eficientes; portanto, quando n_j é grande, o valor da variância assintotica de cada $\hat{\beta}_i$ é mínimo e igual a $1/I_i$, onde:

$$I_i = E\left[\left(\frac{\partial \ell \Phi}{\partial\beta_i}\right)^2\right]$$

, sendo Φ a distribuição conjunta das variáveis amostrais.

Finalmente, a estimativa fornecida é suficiente e, portanto, no conceito de Fisher, extrai toda a informação contida na amostra, informação esta relevante para os parâmetros que estão sendo estimados.

A principal deficiência do método é que ele exige, do mesmo modo que os anteriormente descritos neste trabalho, dados agrupados, ou seja, $n_j > 1$, $j=1,2,...,N$. No entanto, conforme enfatizado por Berkson (1953), o método fornece praticamente as mesmas estimativas e estatísticas de teste que o método de máxima verossimilhança, descrito a seguir. Assim, dispondo-se de dados agrupados e de um grande número de observações, é justificado o uso do método qui-quadrado logit mínimo, uma vez que, nessas condições, seus resultados são obtidos de forma mais simples do que as do método de máxima verossimilhança.

Tendo-se em vista essas dificuldades, torna-se necessário desenvolver métodos que sejam aplicáveis a uma observação por nível e que tenham propriedades aceitáveis para pequenas amostras.

3.4 - O MÉTODO DE MÁXIMA VEROSSIMILHANÇA

Com este método, busca-se eliminar - ou, pelo menos, reduzir - algumas das dificuldades associadas aos métodos anteriores.Aqui, ao invés de estimar a função L_j, trabalha-se diretamente com a função de probabilidade conjunta da amostra. O que se faz é derivar uma expressão para a probabilidade de obter a sequência particular obtida de sucessos ($Y_j=1$) e insucessos ($Y_j=0$). Esta expressão, denominada função de verossimilhança, depende dos parâmetros β, que são desconhecidos; a idéia do método é usar como estimativa de β o valor $\hat{\beta}$ que maximiza essa função.

Seja $n_j=1$, $j=1,2,\ldots,N$. A função de verossimilhança Φ é dada por:

$$\Phi(Y_1,\ldots,Y_N, X_1,\ldots,X_N) = \prod_{j=1}^{N} P_j^{Y_j} (1 - P_j)^{1-Y_j}$$

ou

$$\Phi(Y_1,\ldots,Y_N, X_1,\ldots,X_N)= \prod_{j=1}^{N} \left[\frac{1}{1 + e^{-X_j\beta}}\right]^{Y_j} \left(\frac{e^{-X_j\beta}}{1 + e^{-X_j\beta}}\right)^{1-Y_j}$$

Se $n_j > 1$ para algum j, tem-se:

$$Y_{ij} = \begin{cases} 1 & \text{com probabilidade } P_j \qquad \text{para } i = 1,2,\ldots,n_j \\ 0 & \text{com probabilidade } 1 - P_j \qquad\quad j = 1,2,\ldots,N \end{cases}$$

e Φ é dada por

$$\Phi(Y_{11},\ldots,Y_{n_1,1},\ldots,Y_{1N},\ldots Y_{n_N,N},X_1,\ldots,X_N)=\prod_{j=1}^{N} \prod_{i=1}^{n_j} P_j^{Y_{ij}}(1-P_j)^{1-Y_{ij}}$$

Seja Φ^* o logaritmo natural de Φ, ou seja:

$$\Phi^* = - \sum_{j=1}^{N} \left[(1 - Y_j)X_j\beta + \ell g\ (1 - e^{-X_j\beta})\right]$$

Como Φ^* é uma função monotonicamente crescente de Φ, maximizar uma é equivalente a maximizar a outra. Assim, derivando-se Φ^* com relação a β e igualando-se a derivada a zero, obtém-se o seguinte sistema de equações:

$$\sum_{j=1}^{N} \left[\frac{1}{1 + e^{-X_j\hat{\beta}}}\right] X_j = \sum_{j=1}^{N} Y_j X_j \qquad \text{III.4}$$

As Equações 4 não podem ser resolvidas explicitamente, mas há métodos para a obtenção iterativa de uma série de aproximações que convergem para a solução ótima, se esta existe. Um exame das condições de segunda ordem mostra que esta solução realmente maximiza Φ^*.

A dificuldade básica no uso do método de máxima verossimilhança é seu custo. Para amostras muito grandes, com muitas variáveis, o método é relativamente dispendioso. No entanto, ele tem a vantagem de poder ser empregado quer sejam as variáveis explicativas categóricas, como sexo, raça, religião, etc; quer elas tenham sido previamente categorizadas; ou caso se tenham variáveis contínuas e categóricas, ao mesmo tempo. Além disso, o uso do método não requer agregação a priori das variáveis explicativas, e nem a consequente suposição de que indivíduos com diferentes características têm as mesmas probabilidades de fazer uma dada escolha.

No entanto, as convenientes propriedades possuídas pelos estimadores obtidos são válidas para grandes amostras. Para amostras pequenas, as propriedades do estimador de máxima verossimilhança do modelo logit são desconhecidas, com exceção de alguns casos especiais, como observa McFadden (1974). Uma implementação desse método encontra-se em Nerlove e Press (1973) que usam o método de Newton-Raphson para maximizar Φ^*.

CAPÍTULO IV

CONCLUSÕES

A escolha do método mais apropriado para estimar os parâmetros do modelo logit dependerá dos dados disponíveis e das restrições existentes, principalmente no que diz respeito a custos computacionais. Se houver disponibilidade de dados agrupados, se as variáveis explicativas são categóricas e se a amostra é grande, o método de mínimos quadrados generalizados fornece estimadores satisfatórios. No entanto, se custos computacionais não representam um problema, o método de máxima verossimilhança é desejável, uma vez que ele garante estimativas consistentes dos parâmetros, no caso de grandes amostras.

REFERÊNCIAS BIBLIOGRÁFICAS

BERKSON, J. *A Statiscally Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response, Based on the Logistic Function.* Journal of the American Statistical Association, 48 (263), 565-599, Setembro, 1953.

HANUSHEK, E.A.; JACKSON, J.E. *Statistical Methods for Social Scientists.* New York, Academic Press, 1977.

McFADDEN, D. *Conditional Logit Analysis of Qualitative Choice Behavior,* in P.Zarembka, ed; Frontiers in Econometrics, New York, Academic Press, 1974.

NERLOVE, M.; PRESS, S.J. *Univariate e Multivariate Log-Linear and Logistic Models.* RAND.R1306 - EDA/NIH, Santa Mônica, 1973.

THEIL, H. *On the Estimation of Relationships Involving Qualitative Variables.* American Journal of Sociology, 76(1), 103-154, Julho, 1970.