

Resumo Modelos Lineares Generalizados

Os modelos de regressão são ferramentas que auxiliam no processo de tomada de decisão, já que permitem explicar ou fazer previsões do valor de uma variável aleatória em função de outras. Chamamos a variável a ser explicada de variável resposta, de modo que, dependendo do tipo de sua distribuição, escolhe-se uma técnica de regressão capaz de explicá-la através de covariáveis.

Os **Modelos Lineares Generalizados** foram propostos para aplicações onde a variável de resposta \mathbf{y} pode ser representada por alguma distribuição da família de exponencial, univariada como por exemplo as distribuições normal, binomial, binomial negativa, gama, Poisson e normal inversa.

Sendo uma extensão dos Modelos Lineares simples os MLG envolvem uma variável resposta univariada, variáveis explicativas e uma amostra aleatória de n .

Definição

Os modelos lineares generalizados podem ser usados quando se tem uma única variável aleatória Y e associado a ela um conjunto de variáveis explicativas X_1, \dots, X_p . Para uma amostra de n observações (y_i, \mathbf{x}_i) em que $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})^T$ é o vetor coluna de covariáveis. Sendo os dados para o modelo organizados a seguinte forma

x_1	x_2	...	x_k	y
x_{11}	x_{12}	...	x_{1k}	y_1
x_{21}	x_{22}	...	x_{2k}	y_2
.
.
.
x_{n1}	x_{n2}	...	x_{nk}	y_n

temos que os MLG envolvem 3 componentes:

- i) **Componente Aleatório:** é a variável resposta do modelo que deve ter uma distribuição pertencente à família exponencial na forma canônica.

Sendo representado por um conjunto de variáveis aleatórias independentes Y_i, \dots, Y_k provenientes de uma mesma distribuição que faz parte da família

exponencial canônica com médias μ_1, \dots, μ_k , ou seja

$$E(Y_i) = \mu_i, i = 1, 2, \dots, n \quad (1)$$

um parâmetro constante de escala, conhecido $\phi > 0$ e que depende de um único parâmetro θ_i , chamado parâmetro canônico ou **parâmetro de localização**. A f.d.p. de Y_i é dada por

$$f(y_i; \theta_i, \phi) = \exp \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i; \phi) \right\} \quad (2)$$

sendo $b(\cdot)$ e $c(\cdot)$ funções conhecidas. Em geral, $a_i(\phi) = \frac{\phi}{w_i}$ sendo pesos a priori

ii) **Componente Sistemático:** são as variáveis explicativas, que entram na forma de uma soma linear de seus efeitos.

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j = \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{ou} \quad \boldsymbol{\eta} = \mathbf{X} \boldsymbol{\beta} \quad (3)$$

sendo $\mathbf{X} = (x_1, \dots, x_n)^T$ a matriz do modelo. $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ o vetor de parâmetros e $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ o preditor linear. De modo geral podemos definir o preditor linear como

$$\eta = \beta_0 + \beta_1 x_1, \dots, \beta_k x_k \quad (4)$$

iii) **Função de ligação:** a ligação entre os componentes aleatório e sistemático é feita através de uma função (por exemplo, logarítmica para os modelos log-lineares). Desse modo, uma função que liga o componente aleatório ao componente sistemático, ou seja, relaciona a média ao preditor linear, isto é,

$$\eta_i = g(\mu_i) = g(\beta_0 + \beta_1 x_{i1}, \dots, \beta_k x_{ik}) \quad (5)$$

sendo $g(\cdot)$ uma função monótona (preserva a relação de ordem), derivável.

Precisamos que conjunto o de parâmetros β_1, \dots, β_p sejam uma combinação linear igual a alguma função do valor esperado de Y_i conjunto menor de parâmetros tais que uma combinação linear dos seja igual a alguma função do valor esperado de Y

- Distribuição da variável resposta;
- matriz modelo;

- função de ligação

Se a função de ligação é escolhida de tal forma que $g(\mu) = \theta$ o preditor linear modela diretamente o parâmetro canônico e tal função de ligação é chamada ligação canônica. Assim, as estatísticas dessa ligação são **garantidamente suficientes**. As **funções de ligação canônicas** são apresentadas na tabela abaixo

Distribuição	Ligação canônica
Normal	Identidade: $\eta = \mu$
Poisson	Logarítmica: $\eta = \ln(\mu)$
Binomial	Logística: $\eta = \ln\left(\frac{\pi}{1-\pi}\right) = \ln\left(\frac{\mu}{m-\mu}\right)$
Gama	Recíproca: $\eta = \frac{1}{\mu}$
Normal Inversa	Recíproca ² : $\eta = \frac{1}{\mu^2}$

Para o modelo linear clássico a função de ligação é chamada identidade, pois o preditor linear é igual à média. Essa função de ligação é adequada no sentido em que ambos, η e μ , podem assumir valores na linha real. A seleção da função de ligação em MLG pode ser vista como o equivalente da escolha da transformação da resposta no modelo linear de regressão.

Estimação dos Parâmetros: ajuste do modelo

O ajuste de um modelo linear generalizado é determinado pelo vetor $\hat{\beta}$ de estimativas dos parâmetros. A estimação desses parâmetros é feita através da maximização da função de log-verossimilhança:

$$\ell = \ell(\theta; y) = \sum_{i=1}^n \ln f(y_i; \theta_i, \phi) = \sum_{i=1}^n \left\{ \frac{1}{a_i(\phi)} [y_i \theta_i - b(\theta_i)] + c(y_i; \phi) \right\} \quad (6)$$

Uma propriedade da família exponencial de distribuições é que seus elementos satisfazem a condições de regularidade suficientes para assegurar que o máximo global do logaritmo da função de verossimilhança $\ell(\theta; y)$ é dado pela solução do sistema de

equações $U_\theta = \frac{d\ell}{d\theta} = 0$, ou seja, podemos obter os estimadores de máxima verossimilhança (EMV).

Assim, tem-se, então, que a função escore é dada por

$$U_j = \sum_{i=1}^n \frac{\partial \ell_i}{\partial \beta_j} = \sum_{i=1}^n \frac{1}{a_i(\phi)} (y_i - \mu_i) \frac{1}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} \quad (7)$$

Em geral, as equações $U_j = 0, j = 1, 2, \dots, p$ não são lineares e têm que ser resolvidas numericamente por processos iterativos do tipo Newton-Raphson.

Algoritmo para Estimar os Parâmetros

Quando as derivadas de 2ª ordem são obtidas facilmente, o método de Newton-Raphson é bastante útil. Acontece, porém, que isso nem sempre ocorre e no caso dos modelos lineares generalizados usa-se o método escore de Fisher que, em geral, é mais simples.

Ele envolve a substituição da matriz de derivadas parciais de 2ª ordem pela matriz de valores esperados das derivadas parciais, isto é, a substituição da matriz de informação observada, $J(\theta)$, pela matriz de informação esperada de Fisher $I(\theta)$.

O algoritmo para obter a estimativa de máxima verossimilhança é denominado algoritmo dos mínimos quadrados ponderados iterativo (MQPI). Sendo

$$\beta^{(m+1)} = (X^T W^{(m)} X)^{-1} X^T W^{(m)} z^m \quad (8)$$

em que $W^{(m-1)} = \frac{1}{V(\mu_i^{(m)})} \left(\frac{\partial \mu_i^{(m)}}{\eta_i^{(m)}} \right)$ e $(X^T W X)^{-1}$ é a matriz informação de **fisher**

Então $\beta^{(m+1)}$ que tem a forma da solução das equações normais, para o modelo linear obtida pelo método dos quadrados mínimos ponderados, exceto que nesse caso a solução $\hat{\beta} = \beta^{(m+1)}$ é obtida por processo numérico iterativo.

O método usual para iniciar o processo iterativo é especificar uma estimativa inicial $\beta^{(0)}$ e sucessivamente alterá-la até que a convergência seja obtida e $\hat{\beta} = \beta^{(m+1)}$.

Propriedades e distribuição amostral de $\hat{\beta}$

Na inferencial de MLG, a idéia básica é que se $\hat{\theta}$ é um estimador consistente para um parâmetro θ e $Var(\hat{\theta})$ é a variância desse estimador, então, para amostras grandes, $\hat{\theta}$ assintoticamente imparcial.

Logo, $\hat{\beta}$ segue

$$\hat{\beta} \sim N(\beta, (\phi X^T W X)^{-1}) \quad (9)$$

que é a base para a construção de testes e intervalos de confiança para os modelos lineares generalizados.

Deviance

O ajuste de um modelo a um conjunto de dados observados y pode ser encarado como uma maneira de se substituir y por um conjunto de valores estimados $\hat{\mu}$ para um modelo com um número de parâmetros relativamente pequeno. Logicamente os $\hat{\mu}$'s não serão exatamente iguais aos y 's, e a questão, então, que aparece é em quanto eles diferem. Isto porque, uma discrepância pequena pode ser tolerável enquanto que uma grande, não.

Assim, admitindo-se uma combinação satisfatória da distribuição da variável resposta e da função de ligação, o objetivo é determinar quantos termos são necessários na estrutura linear para uma descrição razoável dos dados. Um número grande de variáveis explanatórias (ou covariáveis) pode levar a um modelo que explique bem os dados mas tenha alta complexidade na interpretação. Por outro lado, um número pequeno de covariáveis pode levar a um modelo de interpretação fácil, porém, que se ajuste pobremente aos dados. O que se deseja na realidade é um modelo intermediário.

O modelo mais simples é o **modelo nulo** que tem um único parâmetro. No outro extremo, está o **modelo saturado** ou completo que tem n parâmetros, um para cada variável. Na prática o modelo nulo é simples demais e o modelo saturado não é informativo, pois não resume os dados, mas simplesmente os repete. Além destes temos os **modelos minimal e maximal**, seguem os exemplos:

Nulo: $\eta_i = \mu$

Minimal: $\eta_i = \mu + \beta_\ell$

Maximal: $\eta_i = \mu + \beta_\ell + \alpha_j + \gamma_k + (\alpha\gamma)_{jk}$

Saturado: $\eta_i = \mu + \beta_\ell + \alpha_j + \gamma_k + (\alpha\gamma)_{jk} + (\beta\alpha)_{\ell j} + (\beta\gamma)_{\ell k} + (\beta\alpha\gamma)_{\ell jk}$

Qualquer modelo com p parâmetros, situado entre os modelos minimal e maximal, pode ser o modelo de interesse. Para selecionar o melhor modelo podemos medir seu ajuste com uma medida de discrepância a **deviance**, dada por:

$$S_p = 2(\hat{\ell}_n - \hat{\ell}_p) \quad (10)$$

sendo $\hat{\ell}_n$ e $\hat{\ell}_p$ as log-verossimilhanças para os modelos saturados e o de interesse, respectivamente. Pode-se, ainda escrever

$$S_p = \frac{1}{\phi} \sum_{i=1}^n d_i^2 \quad (11)$$

sendo que d_i^2 mede a diferença dos logaritmos das funções de verossimilhanças observada e ajustada, chamado componente da deviance. Sendo uma medida da distância dos valores ajustados $\hat{\mu}$'s em relação aos dados observados y 's.

Scaled Deviance

$$S_p = \frac{1}{\phi} \sum_{i=1}^n 2w_i \left\{ y_i \left[\tilde{\theta}_i - \hat{\theta}_i \right] - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right\} \quad (12)$$

Tabela 9: Funções *deviance* para algumas distribuições

Distribuição	<i>Scaled deviance</i>
Normal	$S_p = \frac{1}{\sigma^2} \sum_{i=1}^n w_i (y_i - \hat{\mu}_i)^2$
Binomial	$S_p = 2 \sum_{i=1}^n w_i \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) + (m_i - y_i) \ln \left(\frac{m_i - y_i}{m_i - \hat{\mu}_i} \right) \right]$
Poisson	$S_p = 2 \sum_{i=1}^n w_i \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right]$
Binomial negativa	$S_p = 2 \sum_{i=1}^n w_i \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i + k) \ln \left(\frac{y_i + k}{\hat{\mu}_i + k} \right) \right]$
Gama	$S_p = 2v \sum_{i=1}^n w_i \left[-\ln \left(\frac{y_i}{\hat{\mu}_i} \right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right]$
Normal Inversa	$S_p = \frac{1}{\sigma^2} \sum_{i=1}^n w_i \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2}$

Escolha da componente aleatória e da ligação

- Quando a resposta Y é resultado de uma contagem utilizamos a Poisson com ligação logarítmica.
- Quando temos uma resposta Y dicotômica utilizamos a Bernoulli com a ligação logit.

- Quando temos um total de casos n e precisamos ver a número de sucessos em uma respostas Y que permite a entrada $\frac{p}{n}$ usamos à Binomial.

Resultados da família exponencial canônica

Identificadores da Família Exponencial para Algumas Distribuições						
Distribuição	$a(\phi)$	θ	$b(\theta)$	$c(y; \phi)$	$\mu(\theta)$	$V(\mu) = W_i$
Normal $N(\mu, \sigma^2)$	σ^2	μ	$\frac{\theta^2}{2}$	$-\frac{1}{2} \left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right]$	θ	1
Poisson $P(\mu)$	1	$\ln \mu$	e^θ	$-\ln y!$	e^θ	μ
Binomial $B(m, \pi)$	1	$\ln \left(\frac{\pi}{1-\pi} \right)$	$m \ln(1 + e^\theta)$	$\ln \binom{m}{y}$	$m \frac{e^\theta}{1 + e^\theta}$	$\frac{1}{m} \mu(m - \mu)$
Bin. Negativa $\text{BinNeg}(\mu, k)$	1	$\ln \left(\frac{\mu}{\mu + k} \right)$	$-k \ln(1 - e^\theta)$	$\ln \left[\frac{\Gamma(k + y)}{\Gamma(k) y!} \right]$	$k \frac{e^\theta}{1 - e^\theta}$	$\mu \left(\frac{\mu}{k} + 1 \right)$
Gama $G(\mu, \nu)$	ν^{-1}	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$\nu \ln(\nu y) - \ln y - \ln \Gamma(\nu)$	$-\frac{1}{\theta}$	μ^2
Normal Inversa $IG(\mu, \sigma^2)$	σ^2	$-\frac{1}{2\mu^2}$	$-(-2\theta)^{1/2}$	$-\frac{1}{2} \left[\ln(2\pi\sigma^2 y^3) + \frac{1}{\sigma^2 y} \right]$	$(-2\theta)^{-1/2}$	μ^3

Sendo W_i uma matriz diagonal

Proc Genmod SAS

Funções de ligação e seus defeaut no SAS

DIST=	Distribution	Default Link Function
BINOMIAL BIN B	binomial	logit
GAMMA GAM G	gamma	inverse (power(-1))
GEOMETRIC GEOM	geometric	log
IGAUSSIAN IG	inverse Gaussian	inverse squared (power(-2))
MULTINOMIAL MULT	multinomial	cumulative logit
NEGBIN NB	negative binomial	log
NORMAL NOR N	normal	identity
POISSON POI P	Poisson	log
ZIP		

(13)

Referências

- [1] Demetrio, Clarice Garca Borges. Modelos lineares generalizados em experimentação agronômica. Universidade Nacional de Colombia. Facultade Ciencias. Departamento de Estatística, 2002
- [2] Dobson, Annette J., and Adrian Barnett. An introduction to generalized linear models. CRC press, 2008.