

4

Modelos Lineares Generalizados

Neste capítulo, serão apresentados arcabouços teóricos dos modelos lineares generalizados (MLGs) e como casos particulares desses modelos são aplicáveis ao problema da estimação das perdas não técnicas. É interessante deixar claro que os MLGs são uma extensão dos modelos lineares, e assim incluem o modelo de regressão linear (modelo com a variável resposta seguindo distribuição normal). No fundo, o MLG já é uma classe de modelos utilizada pela maioria dos pesquisadores, entretanto nem todos sabem que estão lidando com um caso particular dos MLG's.

4.1.

Introdução aos MLG's

Segundo Turkiman e Silva (2000), o modelo de regressão linear, desenvolvido por Legendre e Gauss no começo do século XIX, foi a principal técnica de modelagem estatística até meados do século XX, embora vários modelos não lineares ou não normais já tenham sido desenvolvidos em face de situações que não eram adequadamente explicadas pelo modelo linear normal, antes desse período. Podem ser citados como exemplos: modelos de McCullaghand Nelder (1989) e Lindsey (1997), o modelo complementar log-log para ensaios de diluição (Fisher, 1922), os modelos probit (Bliss, 1935), e logit (Berkson, 1944) para proporções, os modelos log-lineares para dados de contagens (Birch,1963), entre outros.

Ainda de acordo com Turkiman e Silva (2000), os modelos supracitados apresentam estrutura de regressão linear e possuem em comum o fato da variável resposta seguir uma distribuição dentro de uma família de distribuições com propriedades muito particulares: a família exponencial. Os modelos lineares generalizados foram introduzidos por Nelder e Wedderburn (1972) e nada mais são do que uma síntese destes e de outros modelos, vindo assim unificar, tanto do ponto de vista teórico como conceitual, a abordagem linear generalizada.

4.2.

Estrutura do Modelo

Conforme discutido anteriormente, pode-se dizer que, para se enquadrar na classe dos modelos lineares generalizados, é necessário que a função de distribuição da variável resposta de determinado problema pertença à família exponencial¹³. De maneira geral, uma distribuição é da família exponencial caso sua função de densidade possa ser descrita sob a seguinte estrutura:

Equação 4 – Estrutura da Família Exponencial

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

Onde y é a variável de interesse e $a(\phi)$, $b(\theta)$ e $c(y, \phi)$ são funções denominadas de funções específicas. O parâmetro θ é o parâmetro de localização e ϕ é o parâmetro de dispersão, associado à variância. Simplesmente calculando-se as derivadas de primeira e segunda ordem da função $b(\theta)$, pode-se obter a média e a variância da variável y , como demonstrado na Equação 5.

Equação 5 – Média e Variância de uma Variável Aleatória da Família Exponencial

$$E(y) = \mu = \frac{db(\theta)}{d\theta} \quad e$$

$$\text{var}(y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi).$$

Como se pode observar, a variância de y é, na verdade, um produto que depende de sua média e de $a(\phi)$. A parcela correspondente à segunda derivada de $b(\theta)$ é conhecida como função da variância, $V(\mu)$. O que faz sentido, dado que em algumas distribuições a variância dos dados muda conforme sua média. É interessante comparar, por exemplo, as distribuições normal (μ , σ) e poisson (λ).

¹³ Família exponencial uniparamétrica.

A variância da distribuição Poisson é dada por λ^{-1} , ou seja, depende diretamente de sua média λ , enquanto a variância da normal, dada por σ , não depende de sua média, mas sim de uma função de variância, que neste caso é igual a 1, como pode ser deduzido a partir da Equação 6.

Equação 6 – Função da variância

$$V(\mu) = \frac{\text{var}(y)}{a(\phi)} = \frac{d^2 b(\theta)}{d\theta^2} = \frac{d\mu}{d\theta}$$

É interessante observar como as principais distribuições de probabilidade são escritas sob a forma estrutural da família exponencial. Na tabela abaixo, é apresentada a associação usual entre a distribuição e o tipo de dado que ela geralmente modela:

Tabela 3 – Algumas distribuições e tipos de dados

Distribuição	Tipos de Dados
Poisson	Contagens
Binomial	Proporções
Normal	Contínuos Simétricos
Gama	Contínuos Assimétricos
Normal Inversa	Contínuos Assimétricos

Sabe-se que a variável resposta “perdas não técnicas” é contínua e assimétrica. Logo, analisando o histograma da mesma junto das informações disponíveis na Tabela 3, percebe-se que as distribuições gama e normal inversa são potencialmente boas candidatas a componente aleatória do modelo linear generalizado, uma vez que são usadas para modelar dados contínuos e assimétricos. Sendo assim, função densidade de probabilidade e as funções $a(\phi)$, $b(\theta)$ e $c(y, \phi)$ serão detalhadas apenas para essas duas distribuições.

a) Gama

$$\begin{aligned}
 f(y; \mu, \alpha) &= \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha}{\mu} \right)^\alpha y^{\alpha-1} e^{-\alpha(y/\mu)} \\
 &= \exp \left[\alpha \ln \alpha - \alpha \ln \hat{\mu} + (\alpha - 1) \ln y_i - \frac{\alpha y}{\mu} \ln \Gamma(\alpha) \right] \\
 &= \exp \left[\frac{-y \frac{1}{\mu} - \ln(\mu)}{\frac{1}{\alpha}} + \alpha \ln(\alpha) + (\alpha - 1) \ln(y) - \frac{\alpha y}{\mu} - \ln \Gamma(\alpha) \right]
 \end{aligned}$$

Então, as funções específicas do modelo ficam definidas como:

$$\begin{aligned}
 \theta &= -\frac{1}{\mu}; \quad b(\theta) = -\ln(\mu); \quad a(\phi) = \phi; \quad \phi = \frac{1}{\alpha}; \\
 c(y, \phi) &= \alpha \ln(\alpha) + (\alpha - 1) \ln(y) - \frac{\alpha y}{\mu} - \ln \Gamma(\alpha)
 \end{aligned}$$

Com isso, a média e variância de y são definidas por:

$$E(y) = \frac{db(\theta)}{d\theta} = \mu; \quad \text{var}(y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \mu^2 \phi = \frac{\mu^2}{\alpha}$$

b) Normal Inversa

$$\begin{aligned}
 f(y; \mu, \sigma^2) &= -\frac{1}{\sqrt{2\pi\sigma^2 y^3}} \exp \left[\frac{-(y - \hat{\mu})^2}{2\sigma^2 \hat{\mu}^2 y} \right] \\
 &= \exp \left[\frac{-y \frac{1}{2\mu^2} + \frac{1}{\mu}}{\sigma^2} - \frac{1}{2y\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2 y^3) \right]
 \end{aligned}$$

Então, as funções específicas do modelo ficam definidas como:

$$\theta = -\frac{1}{2\mu^2}; \quad b(\theta) = \frac{1}{\mu}; \quad a(\phi) = \phi; \quad \phi = \sigma^2; \quad c(y, \phi) = -\frac{1}{2y\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2 y^3)$$

Com isso, a média e variância de y são definidas por:

$$E(y) = \frac{db(\theta)}{d\theta} = \mu; \quad \text{var}(y) = \frac{d^2b(\theta)}{d\theta^2} a(\phi) = \mu^3 \phi = \mu^3 \sigma^2$$

4.2.1.

Componentes do Modelo

Uma vez definido como se estruturam as distribuições segundo a família exponencial, deve-se destinar atenção aos componentes do modelo linear generalizado. Abaixo são descritas as três componentes que formam um MLG.

- a) **Componente Aleatória** - seja y um vetor de observações de tamanho n , denotando uma realização de um vetor de variáveis aleatórias Y , independentemente distribuídas com média μ . A distribuição de Y pode ser qualquer uma da família exponencial;
- b) **Preditor Linear** - a parte sistemática do modelo é dada por $\eta = X\beta$, onde $\eta_i = X_i\beta$ é a componente relativa à i -ésima observação; e,
- c) **Função de Ligação (g)** - através desta função, faz-se a conexão entre a média das observações e a parte sistemática, como descrito na equação a seguir.

Equação 7 – Função de Ligação

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

Fazendo um paralelo entre o MLG e o modelo linear, pode-se dizer que a função de ligação teria papel similar a uma transformação na resposta do modelo linear de regressão. Com tudo, deve ter em mente que a função de ligação transforma μ_i , a média de y_i , e não a resposta, o que é uma grande vantagem já que

possibilita análise direta das estimativas do modelo, evitando a necessidade de utilizar a transformação inversa (que em muitos casos são bastante complicadas) nos valores estimados.

4.3.

Estimação

Uma vez já apresentada e discutida a estruturação do modelo através da família exponencial e dos componentes do MLG, pode-se partir para algo mais prático, como, por exemplo, a estimação dos coeficientes do modelo. Nos MLG's utiliza-se a maximização da função de log-verossimilhança, descrita na Equação 8, para se obter as estimativas dos coeficientes.

Equação 8 – Função da de log-verossimilhança

$$L = \ln l(\mathbf{y}; \boldsymbol{\theta}, \phi) = \sum_{i=1}^n \ln f(y_i, \theta_i, \phi) = \sum_{i=1}^n \left\{ \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi) \right\}$$

Entretanto, a maximização dessa função depende da utilização de métodos numéricos. Para resolver este problema, Cordeiro e Demétrio (2008), indicam o processo de otimização iterativo chamado de algoritmo de Newton-Raphson, definido segundo a Equação 9.

Equação 9 – Coeficientes de um MLG via Newton-Raphson

$$\hat{\boldsymbol{\beta}}^{(m+1)} = (\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}^{(m)}\mathbf{z}^{(m)}$$

Onde:

M = Passo iterativo;

$\boldsymbol{\beta}^{(m+1)}$ = Estimativa do vetor de parâmetros na iteração m;

X = Matriz dos valores das variáveis de regressão;

W = Matriz dos pesos; e

z = Vetor das variáveis de ajuste na m-ésima iteração.

O algoritmo iterativo de Newton-Raphson¹⁴, utiliza mínimos quadrados ponderados, onde W é a matriz de pesos que busca garantir que os resíduos do modelo tenham variância constante, ou seja, que o modelo possua resíduos homocedásticos. Resumidamente, pode-se dizer que o algoritmo inicia o processo especificando uma estimativa inicial e vai sucessivamente alterando-a até que a diferença entre o β na iteração $(m + 1)$ e a estimativa anterior seja menor que um ϵ pré-definido, sendo obtida assim a convergência na matriz dos coeficientes estimados.

4.3.1.

Método Stepwise

O passo prático, que é a estimação dos coeficientes do modelo, já foi apresentado. Agora, deseja-se abordar um procedimento que orienta a seleção das variáveis para o modelo e pode ser considerado como um método complementar a estimação não só de modelos lineares generalizados, mas também de várias classes de modelos. Dentre os mais conhecidos métodos de seleção de variáveis, foi escolhido para aplicação neste estudo, o método stepwise. Este se caracteriza por avaliar não só se determinada variável deve entrar no modelo, mas também avaliar se com a inclusão de uma nova variável, aquela ou aquelas já presentes, perdem significância e devam sair do modelo.

Hair (2005), descreve o stepwise como um método de seleção no qual as variáveis são selecionadas para entrar no modelo com base em sua contribuição incremental sobre as variáveis já presentes na equação. Como já dito em linhas gerais, o processo de estimação via stepwise tem basicamente duas etapas: a inclusão e eliminação de variáveis. Essas etapas serão detalhadas a seguir.

A primeira, inclusão, testa as variáveis uma a uma, incluindo-as no modelo caso elas se apresentem estatisticamente significativas. Intercalada a inclusão de uma nova variável há a segunda etapa, eliminação. Nesta, todas as variáveis presentes no modelo são testadas para caso alguma tenha perdido a significância estatística mínima, seja eliminada. A sequência seguida pelo método stepwise é a seguinte:

¹⁴ Para maiores detalhes ver Demétrio e Cordeiro (2008).

- 1) Inicia-se com o modelo vazio;
- 2) A variável mais significativa estatisticamente é incluída no modelo;
- 3) A segunda variável mais significativa estatisticamente é incluída no modelo;
- 4) As variáveis presentes no modelo são testadas e caso alguma tenha perdido significância mínima é excluída do modelo.
- 5) O processo de testes de inclusão e exclusão continua até que não haja mais variáveis significantes para entrar o modelo.

Para determinar se haverá inclusão ou exclusão do modelo, as variáveis são testadas com a estatística t_0 que segue distribuição t-student. Um p-valor *default* foi pré-fixado neste estudo com o valor de 10%. Caso os testes resultantes da estatística t_0 , Equação 10, apontem que o p-valor fique acima do valor base, a variável testada deve ficar fora do modelo final.

Equação 10 – Estatística de teste t-student

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}}$$

Onde:

$\hat{\beta}_j$ = Coeficientes estimados;

4.4.

Pós-estimação

Após a estimação dos coeficientes do modelo, alguns procedimentos devem ser realizados a fim de medir a qualidade do ajuste e adequabilidade do modelo, ou seja, deve ser feita uma validação dos resultados. Num primeiro momento essas análises servem também para comparar diferentes modelos, o que é de interesse neste estudo uma vez que já foi constatado que tanto a distribuição

gama quanto a inversa normal são candidatas à distribuição da variável de interesse.

A fim de escolher qual modelo é o melhor para explicar “perdas não técnicas” estimam-se os modelos candidatos e posteriormente, analisa-se qual possui maior poder explicativo e menor erro, sempre estando de acordo com as premissas dos modelos de regressão, ou seja, se seus resíduos se comportam de maneira esperada¹⁵. Para decidir entre um ou outro modelo, inicialmente propõe-se analisar medidas associadas ao ajuste do modelo com o coeficiente de determinação (R^2), e os critérios de informação (AIC e BIC).

4.4.1.

Coeficiente de Determinação: R^2 e Pseudo R^2

O Coeficiente de determinação, também chamado de R^2 é uma medida de qualidade do modelo econométrico em relação à sua habilidade de estimar corretamente os valores da variável resposta y , seu cálculo é dado através da razão entre a soma dos quadrados explicados e soma dos quadrados totais. O R^2 indica quanto da variância da variável resposta é explicada pela variância das variáveis explicativas. Seu valor está no intervalo de 0 a 1: Quanto maior, mais explicativo é o modelo. Entretanto, para modelos estatísticos baseados em estimativas da máxima verossimilhança, como um MLG, é mais conveniente construir uma estimativa do R^2 , chamada de “pseudo- R^2 ”. Na Equação 11 é descrita a formulação do R^2 .

Equação 11 – R^2

$$R^2 = \frac{SQ_{exp}}{SQ_{tot}} = 1 - \frac{SQ_{res}}{SQ_{tot}}$$

Onde:

SQ_{exp} = Soma dos quadrados explicados;

SQ_{tot} = Soma dos quadrados totais; e

¹⁵ Maiores detalhes em Gujarati (2006).

SQ_{res} = Soma dos quadrados dos resíduos.

No caso do modelo linear generalizado, utiliza-se frequentemente o chamado “ R^2 de McFadden”, essa estatística que tenta reproduzir o R^2 e pode ser calculada pela razão de log-verossimilhança do modelo final com o modelo composto apenas do intercepto (modelo nulo). Sua interpretação é análoga ao R^2 dos modelos de regressão linear. Seus valores situam-se entre 0 e 1 e, quanto mais próximo de 1, melhor o ajuste. O pseudo R^2 de McFadden é descrito segundo a equação abaixo.

Equação 12 – R^2 de McFadden

$$R_{MF}^2 = 1 - \frac{L_1}{L_0}$$

Onde:

L_0 = Log-verossimilhança do modelo nulo; e

L_1 = Log-verossimilhança do modelo final.

4.4.2.

Critérios de Informação

Os critérios de informação são frequentemente utilizados para orientar a seleção dos modelos e objetivam prover uma medida de informação que equilibre uma boa medida de ajustamento e uma especificação reduzida do modelo (parcimônia quanto ao número de parâmetros). Dada sua natureza de penalidade, os melhores modelos são aqueles que apresentam menores critérios de informação.

O critério de informação de Akaike, conhecido simplesmente como AIC, e o critério de informação bayesiano, chamado de BIC, não são testes de hipótese, eles são ferramentas para seleção de modelos. Os critérios de informação apresentam duas vantagens sobre outros métodos de seleção como o R^2 de McFadden: O AIC e o BIC se baseiam na log-verossimilhança como o outro método e, além disso, consideram e penalizam a complexidade do modelo

(número de parâmetros), o que busca o alinhamento com o princípio da parcimônia, essencial na modelagem estatística. Além disso, os modelos que estão sendo comparados não necessitam ser aninhados¹⁶, o que é a principal restrição na comparação de modelos via teste de hipótese. Nas equações subsequentes os critérios AIC e BIC são expressos respectivamente.

Equação 13 – Critério de Informação AIC

$$AIC = \frac{(-2l)}{n} + \frac{(-2k)}{n}$$

Onde:

k = Número de restrições;

n = Tamanho da amostra; e

l = Função de verossimilhança.

Equação 14 – Critério de Informação BIC

$$BIC = \frac{(-2l)}{n} + \frac{(k \log n)}{n}$$

Onde:

k = Número de restrições;

n = Tamanho da amostra; e

l = Função de verossimilhança.

4.4.3.

Resíduos

Para a análise de resíduos num modelo linear generalizado geralmente se opta por utilizar alguma das três definições de resíduos generalizados: Pearson, Anscombe ou Deviance, Cordeiro e Demétrio (2008). Os três tipos são casos especiais da definição de resíduos dada por Cox e Snell (1968). Eles são uma

¹⁶ Modelo da forma reduzida, necessariamente com a mesma estrutura do modelo completo.

função de y e μ , que tem por objetivo tornar a distribuição dos resíduos mais simétrica.

Além obter melhores resultados transformando a função de distribuição dos resíduos em uma função simétrica, os resíduos de Anscombe ainda possuem a propriedade de estabilizar a variância para o caso de modelos com distribuições diferentes da normal. Sendo assim, serão utilizados neste trabalho, primordialmente, os resíduos de Anscombe. O diagnóstico necessário é o mesmo feito em modelo clássico de regressão, ou seja, os gráficos e testes utilizados ao longo deste estudo serão, basicamente, os mesmos que seriam feitos no caso de um modelo de regressão linear como testes e gráficos para detectar possível heterocedasticidade, confirmar se a distribuição dos resíduos é normal, entre outros. Nas equações a seguir são apresentadas as expressões dos resíduos de Anscombe para distribuições gama e normal inversa.

Equação 15 – Resíduos de Anscombe para um modelo gama

$$\frac{3 \left(y^{\frac{1}{3}} - \hat{\mu}^{\frac{1}{3}} \right)}{\hat{\mu}^{\frac{1}{3}}}$$

Equação 16 – Resíduos de Anscombe para um modelo normal inversa

$$\frac{(\log y - \log \hat{\mu})}{\hat{\mu}^{\frac{1}{2}}}$$