Funções de ligação paramétricas.

Fernando Lucambio

Junho 2009

Departamento de Estatística
Universidade Federal do Paraná
Curitiba/PR, 81531–990, Brasil

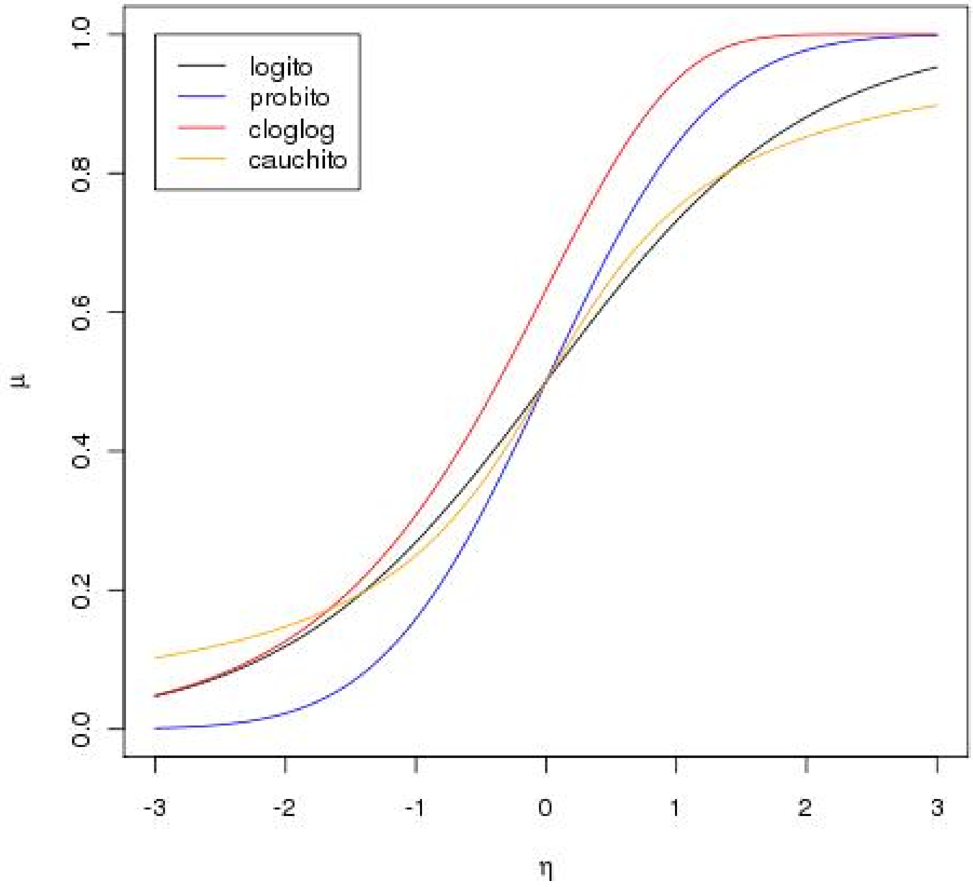
email: lucambio@ufpr.br

Introdução

As funções de ligação logito, probito, cauchito e complementar log-log são de longe as mais utlizadas quando o modelo logístico é o modelo de regressão adequado. Duas propriedades importantes das funções de ligação logito, probito e cauchito é que, quando $\eta = 0$ sempre $\mu = 0.5$ e também de serem simétricas ao redor do ponto $\eta = 0$, isto é, $g^{-1}(a) = g^{-1}(-a)$ para qualquer número real positovo a.

Surgem então questões do tipo, sempre terá que ser assim? ou podemos escolher funções de ligação que não satisfazam estas propriedades?, e desta forma melhorar o ajuste.

Diferentes funções de ligação



Uma alternativa é a função de ligação complemento log-log, a qual não satisfaz nenhuma das propriedades mencionadas.

Mesmo assim, é somente uma alternativa. Estudaremos aqui diferentes alternativas de funções de ligação que não são simétricas e nas quais quando $\eta = 0$ não necessariamente $\mu = 0.5$.

Observemos que no caso da função de ligação complemente log-log se $\eta = 0$ então $\mu = 0.63$. Isto significa que nossas alternativas para valores estimados da média, quando o preditor linear é zero, são somente 0.5 e 0.63.

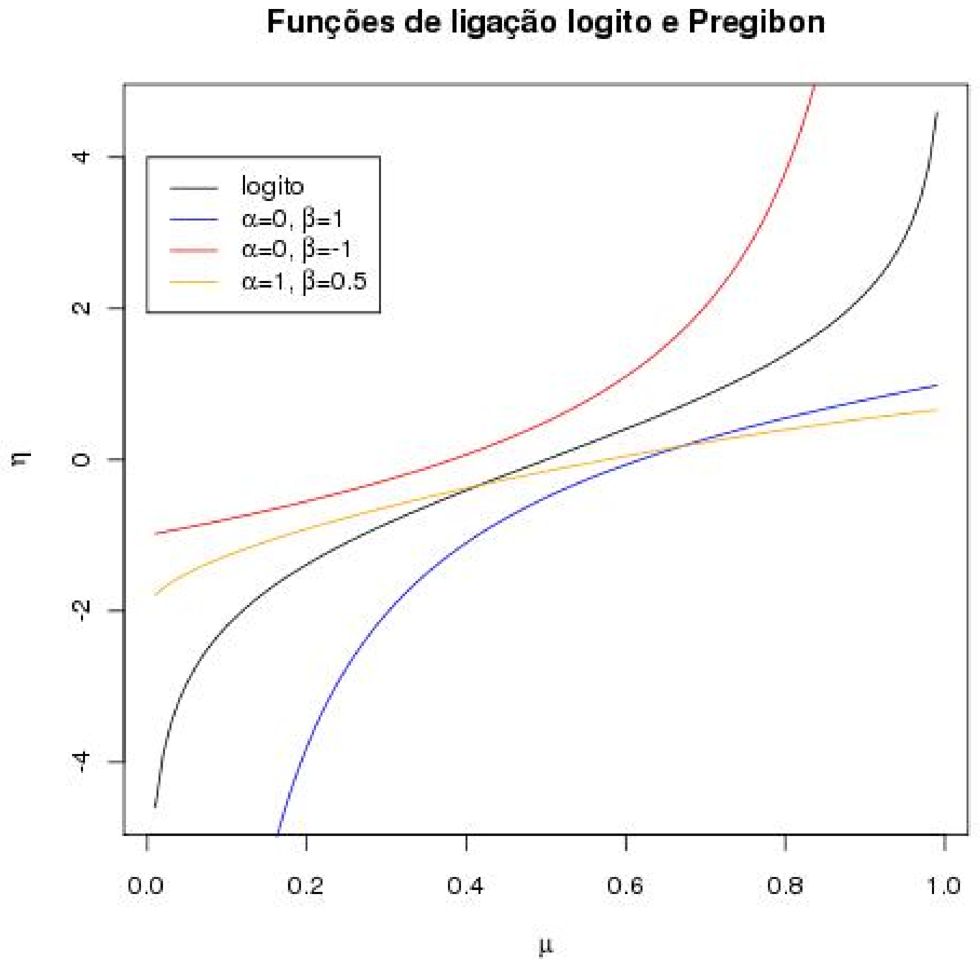
Família de funções de ligação Pregibon

Porposta por Daryl Pregibon em 1980 como generalização à função de ligação logito.

Define-se como:

$$g(\mu; \alpha, \beta) = \frac{\mu^{\alpha-\beta} - 1}{\alpha - \beta} - \frac{(1-\mu)^{\alpha+\beta} - 1}{\alpha + \beta}.$$

Quando $\alpha = \beta = 0$ se reduz a função de ligação logito. Em geral é assimétrica.


Funções de ligação logito e Pregibon
logito
α=0, β=1
α=0, β=-1
α=1, β=0.5
η
μ
4
2
0
-2
-4
0.0
0.2
0.4
0.6
0.8
1.0

Família de funções de ligação Stukel

Proposta por Thérèse A. Stukel em 1988 também como generalização à função de ligação logito, mas preservando a propriedade de que se $\eta = 0$ então $\mu = 0.5$.

Define-se como:

$$\log\left(\frac{\mu}{1-\mu}\right) = h_\alpha(\eta),$$

onde $\alpha = (\alpha_1, \alpha_2)$ é o vetor de parámetros que definem a forma da função de ligação.

Desta forma o modelo proposto é

$$\mu_\alpha(\eta) = \frac{\exp(h_\alpha(\eta))}{1 + \exp(h_\alpha(\eta))}.$$

A função $h_\alpha(\eta)$ é definida como

$$h_\alpha = \begin{cases} \alpha_1^{-1}(\exp(\alpha_1|\eta|) - 1), & \text{se } \alpha_1 > 0 \\ \eta, & \text{se } \alpha_1 = 0 \\ -\alpha_1^{-1}\log(1 - \alpha_1|\eta|), & \text{se } \alpha_1 < 0 \end{cases}$$

se $\eta \geq 0$ ou equivalentemente se $\mu \geq 0.5$, e

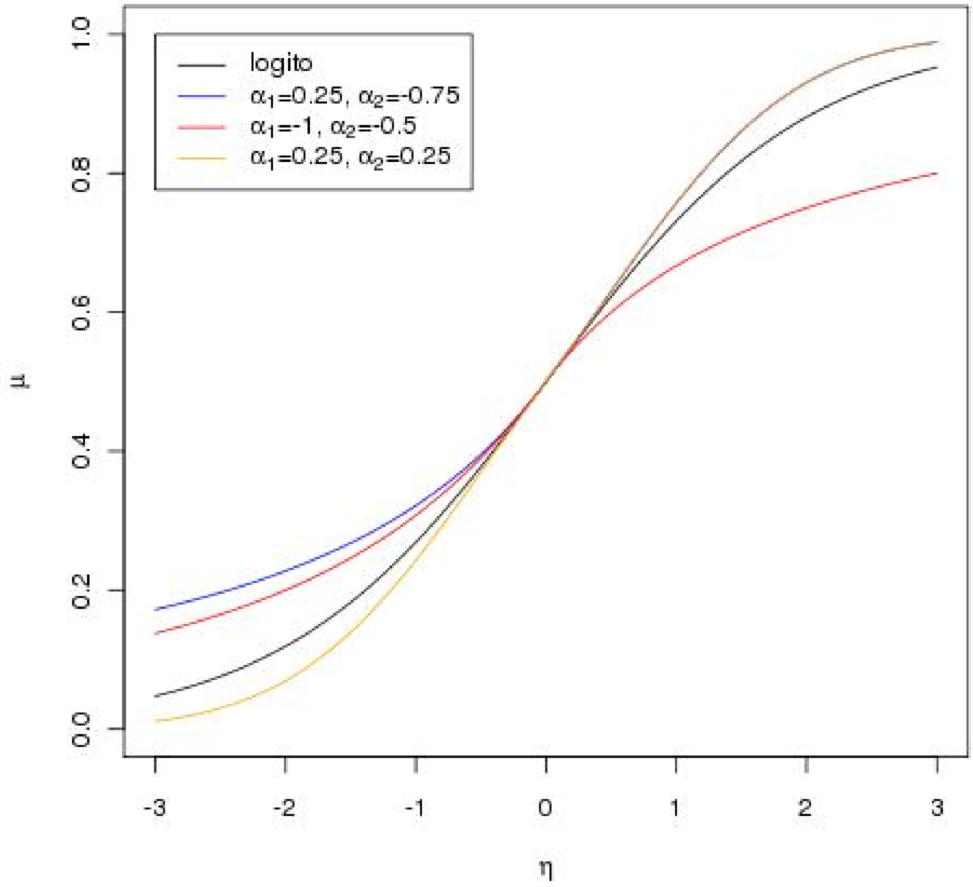
$$h_\alpha = \begin{cases} -\alpha_2^{-1}(\exp(\alpha_2|\eta|) - 1), & \text{se } \alpha_2 > 0 \\ \eta, & \text{se } \alpha_2 = 0 \\ \alpha_2^{-1}\log(1 - \alpha_2|\eta|), & \text{se } \alpha_2 < 0 \end{cases}$$

se $\eta \leq 0$ ou equivalentemente se $\mu \leq 0.5$.

Quando $\alpha_1 = \alpha_2 = 0$ obtemos o modelo logístico. Caso contrário cada parâmetro governa o comportamento da curva de maneira diferente. Se $\alpha_1 = \alpha_2$ a curva obtida é simétrica, se $\alpha_1 \neq \alpha_2$ a função de ligação é assimetrica.

O modelo probito é obtido quando $\alpha_1 = \alpha_2 \approx 0.165$, quando $\alpha_1 = -0.037$ e $\alpha_2 = 0.62$ se obtém a chamada função de ligação log-log e quando $\alpha_1 = 0.62$ e $\alpha_2 = -0.037$ obtem-se a função de ligação complemetar log-log. No caso de $\alpha_1 = \alpha_2 \approx -0.077$ a função de ligação chama-se Laplace.

Funções de ligação Stukel



Verificação da adequação da função de ligação

Suponhamos ajustamos um modelo com a função de ligação $g_1(\mu)$ quando a função de ligação correta, mas desconhecida, é $g_0(\mu)$. A verificação da adequação da função de ligação pode ser realizada supondo que a função de ligação utilizada pertence à mesma família de funções na qual assume-se que $g_0(\mu)$ pertence. Isto supõe a utilização de diversos parâmetros, aqui consideraremos somente dois α e δ.

Definamos:

$$\begin{array}{ll} \text{Função de ligação correta:} & g_0(\mu) = g(\mu; \alpha_0, \delta_0) \\ \text{Função de ligação proposta:} & g_1(\mu) = g(\mu; \alpha_1, \delta_1). \end{array}$$

Utilizando a expansão em série de Taylor de primeira órdem, no ponto $g_1(\mu)$, temos a seguinte relação aproximada

$$g_0(\mu) \cong g_1(\mu) + (\alpha_0 - \alpha_1)D_\alpha(g_1) + (\delta_0 - \delta_1)D_\delta(g_1),$$

onde

$$D_\alpha(g_1) = \left\{\frac{\partial}{\partial\alpha} g(\mu;\alpha,\delta)\right\}\Bigg|_{\alpha=\alpha_1,\delta=\delta_1},$$

e similarmente para $D_\delta(g_1)$. Observemos que agora podemos aproximar a função de ligação correta $g_0(\mu) = \mathbf{x}\beta$ por

$$g_1(\mu) = \mathbf{x}\beta + \mathbf{z}\gamma,$$

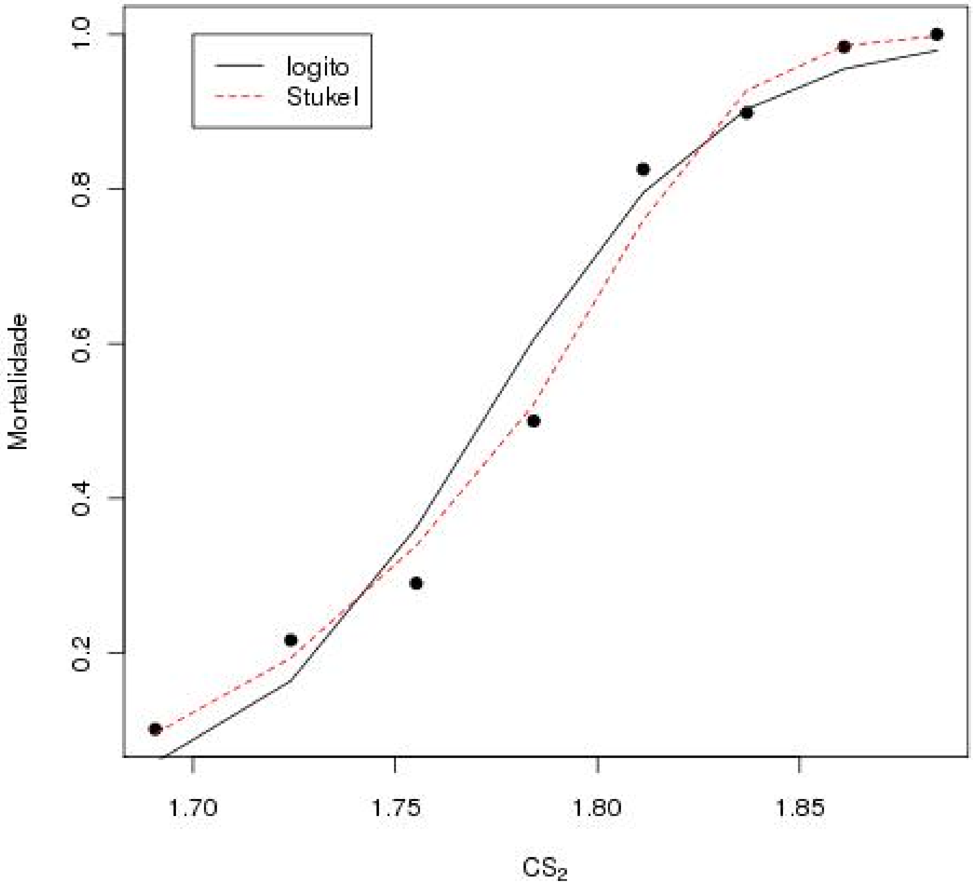
onde $\mathbf{z} = (D_\alpha(g_1), D_\delta(g_1))$ e $\gamma = -(\alpha_0 - \alpha_1, \delta_0 - \delta_1)$.

Nas famílias de funções mencionadas (Pregibon e Stukel) a função de ligação logito se obtém quando ambos parâmetros são zero.

Exemplo

Estudo acerca da mortalidade de escaravelhos adultos depois de cinco horas de exposição a disulfeto de carbono.

CS_2	Mortos	Expostos
1.6907	6	59
1.7242	13	60
1.7552	18	62
1.7842	28	56
1.8113	52	63
1.8369	53	59
1.8610	61	62
1.8839	60	60



logito
Stukel
Mortalidade
0.2
0.4
0.6
0.8
1.0
1.70
1.75
1.80
1.85
CS2

A forma de aproximar as estimativas dos parâmetros do modelo ampliado é escrevendo como preditor linear

$$\eta = \beta_0 + \beta_1 \text{CS}_2 + \frac{1}{2}\alpha_1 \widehat{\eta}^2 \text{I}_{\{\widehat{\eta}>0\}} - \frac{1}{2}\alpha_2 \widehat{\eta}^2 \text{I}_{\{\widehat{\eta}<0\}}.$$

Neste modelo ampliado as estimativas dos coeficientes α_1 e α_2 representam aproximações de primeira ordem das estimativas destes mesmos parâmetros se utilizados diretamente na definição do modelo de regressão.

Este modelo permite além de encontrar aproximações dos parâmetros que definem a nova função de ligação permite também testar se devemos ou não utilizar funções de ligação paramétricas.

Principais Referências

- Pregibon, D. (1980). Goodness of link test for generalized linear models. Applied Statistics, 29, 15-24.

- Stukel, T.A. (1988). Generalized logistic models. Journal of the American Statistical Association, 83, 426-431.