Modelos com distribuições binomial e Poisson e variáveis preditoras

- Até agora os parâmetros das distribuições binomial e Poisson tem sido constantes.

- Binomial: $y \sim Bin(y|N,p)$

- Poisson: $y \sim Po(y|\lambda)$

Modelos com distribuição discreta e variáveis preditoras

- Porém, os parâmetros das distribuições podem depender de funções com variáveis preditoras.

- Binomial: y~Bin (y|N,p)

- y_i| N, p_i ~ Bin (N, p_i)

  logit (p_i) = b_0 + b_1x_1,

- logit (p) = ln(p/1-p)

Modelos discretos com variáveis preditoras

- A função logit lineariza uma curva logística.
- Função da curva logística e relação com a logit.

$$y = e^{a+bx}/(1 + e^{a+bx})$$

$$\left(1 + e^{a+bx}\right) y = e^{a+bx}$$

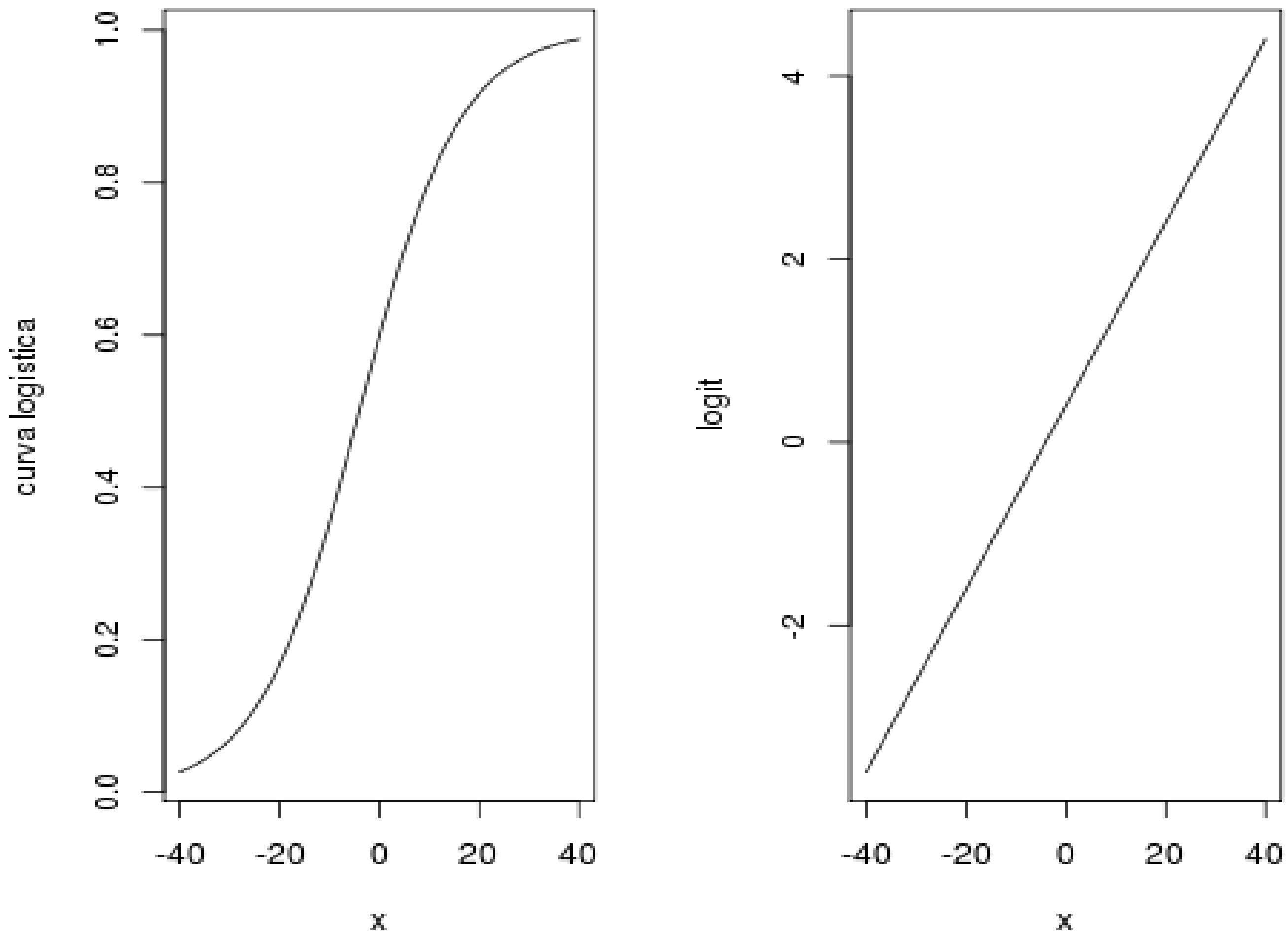
$$y = e^{a+bx}(1 - y)$$

$$\frac{y}{1 - y} = e^{a+bx}$$

$$\log\left(\frac{y}{1 - y}\right) = a + bx$$

Modelos discretos com variáveis preditoras

- A função logit lineariza uma curva logística.



Modelos discretos com variáveis preditoras

- A curva logística tem a propriedade de que qualquer valor que x tome desde -Inf até + Inf, y seguirá restringido entre os valores de 0 a 1.

- Os modelos de distribuição binaria tradicionalmente tem sido analisados com uma regressão logística ou como modelos lineares generalizados de distribuição binomial (GLM), com função de ligação logit.

- Os GLMs tem trés componentes: Um sistemático, outro aleatório e a função de ligação que os vincula.

Modelos lineares generalizados de distribuição binomial

- Parte deterministica, preditor linear: $\boldsymbol{\eta} = \boldsymbol{X\beta}$

  X: Intercepto e variáveis preditoras

- *β:* Coeficiente do intercepto e os coeficientes das variáveis preditoras

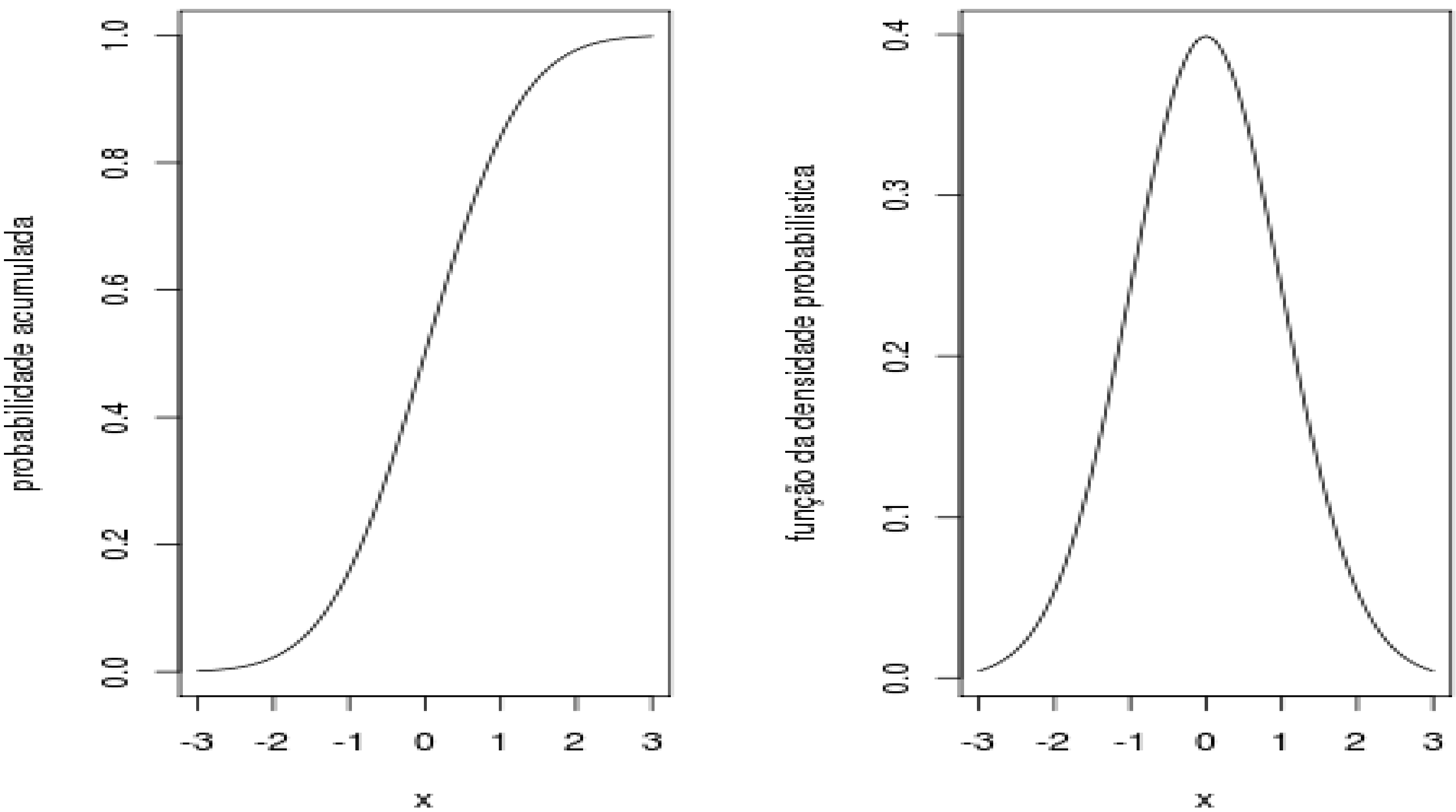
- Parte aléatoria: $y_i \,|\, N, p_i \sim \text{Bin}(N, p_i)$

- Função de ligação $\eta_i = \text{logit}(p)$

  $\log(p_i/1 - p_i) = b_0 + b_1 x_1$,

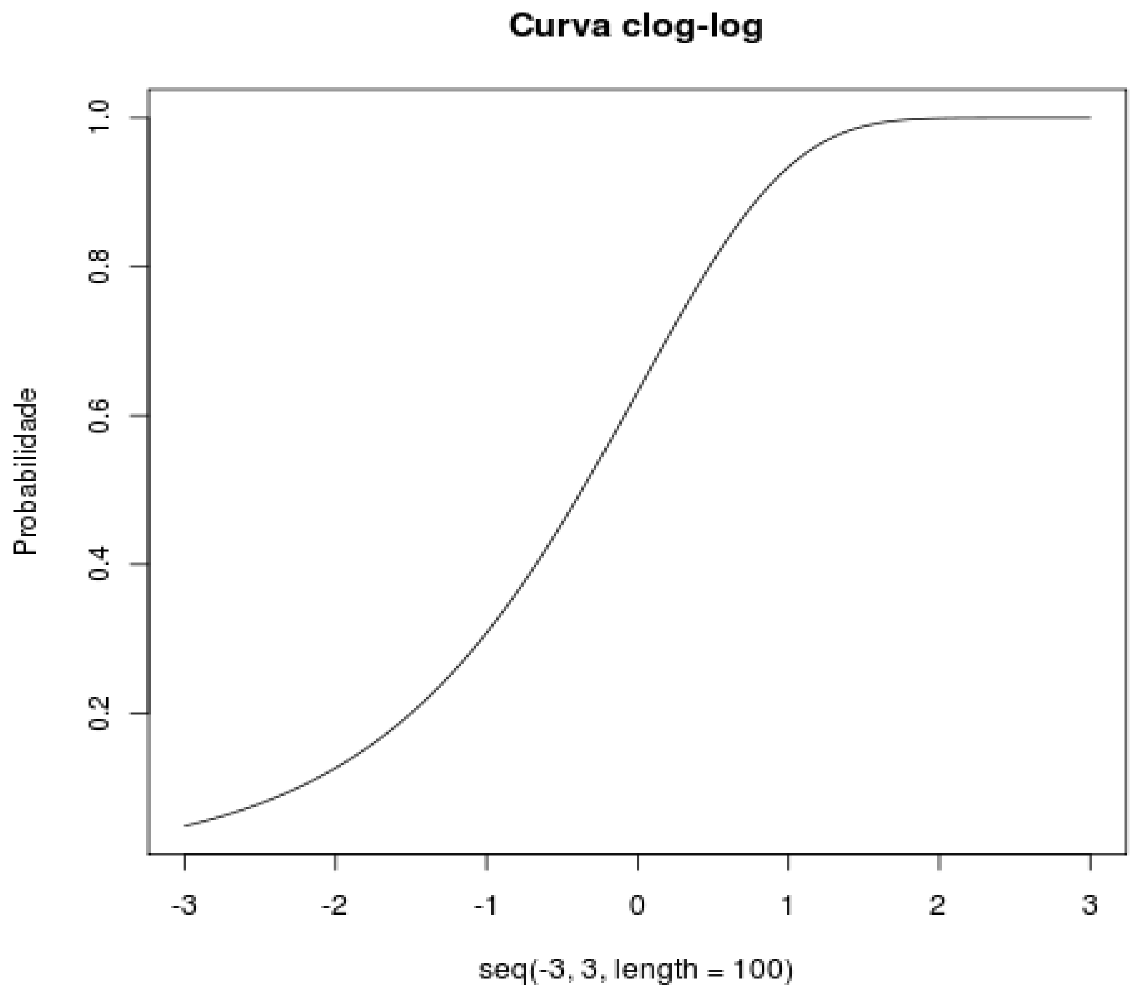
Modelos lineares generalizados de distribuição binomial

Função de ligação probit $\eta = \Phi^{-1}(p) = \boldsymbol{X\beta}$



Modelos lineares generalizados de distribuição binomial,ligação clog-log

- `complog<-function(x) (1-exp(-exp(x)))`



Modelos lineares generalizados de distribuição binomial

Exemplo 1:

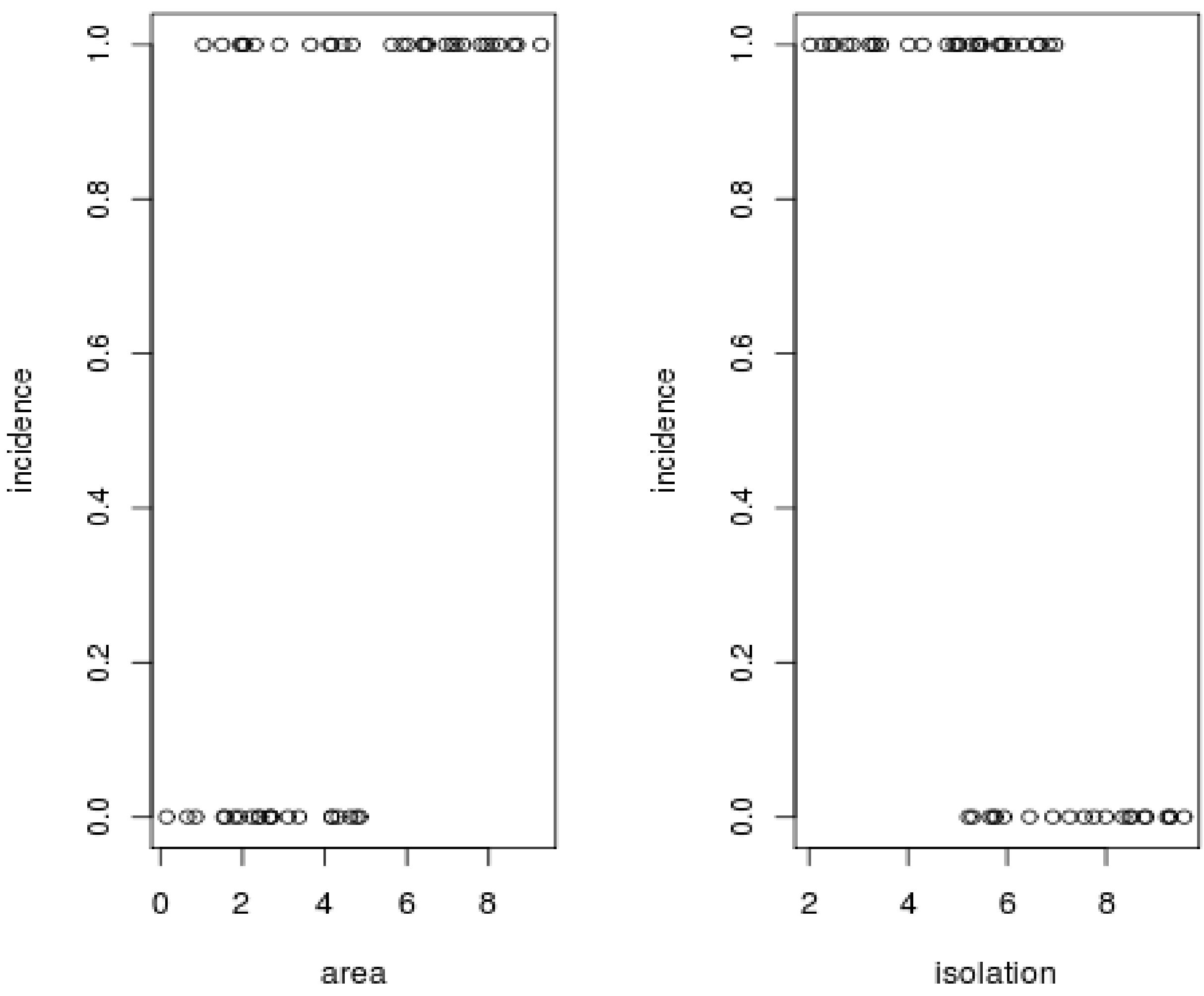
- Reprodução de uma espécie de ave ilhas dependendo da distancia do continente e da área da ilha.

- Exemplo de livro de texto Crawley, M. 2007. The R Book. Wiley, UK.

- Modelos com funções de ligação logit e probit.

Modelos lineares generalizados de distribuição binomial

- Exemplo:Ilhas



Modelos lineares generalizados de distribuição binomial, ligação logit

- Função logística:

- `logistica<-function (x){`

  ```
  exp(x)/(1+ exp(x))
  }
  ```

- Função de máxima verossimilhança
- `chalmod.lg2<-mle2(incidence~dbinom(logistica(a+ b1*area+b2*isolation), size=1), start=list(a =1, b1= 0.05, b2 = -0.13))`

Modelos lineares generalizados de distribuição binomial

Optimizando

- `chalmod.lg1<-mle2(binomLogitNLL, start=list(a=1,b1=0.05,b2=-0.13))`

Alternativamente fazemos os dois passos simultaneamente com a interface mle2

- `chalmod.lg2<-mle2(incidence~dbinom(logistica(a+ b1*area+b2*isolation), size=1), start=list(a =1, b1= 0.05, b2 = -0.13))`

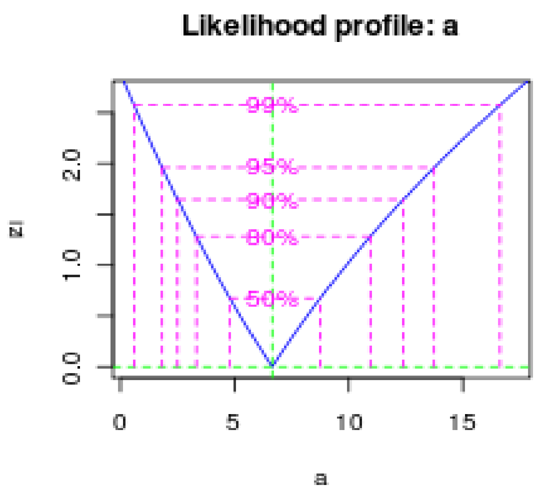
Equivalente a:

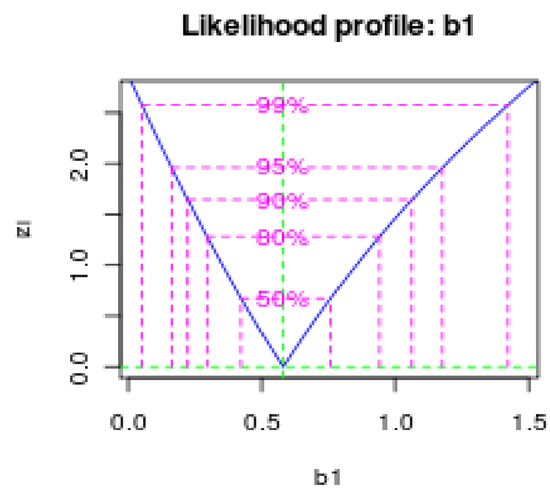
- `chalmod.lg3<-glm(incidence~area+isolation,`

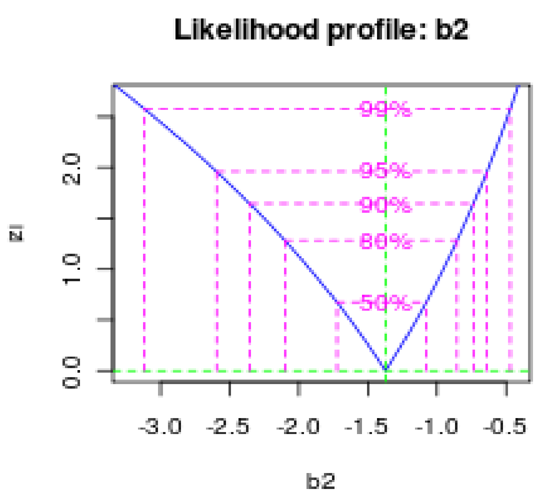
  `binomial(link="logit")`

Modelos lineares generalizados de distribuição binomial

- Perfis de verossimilhança, premissas I. confiança.

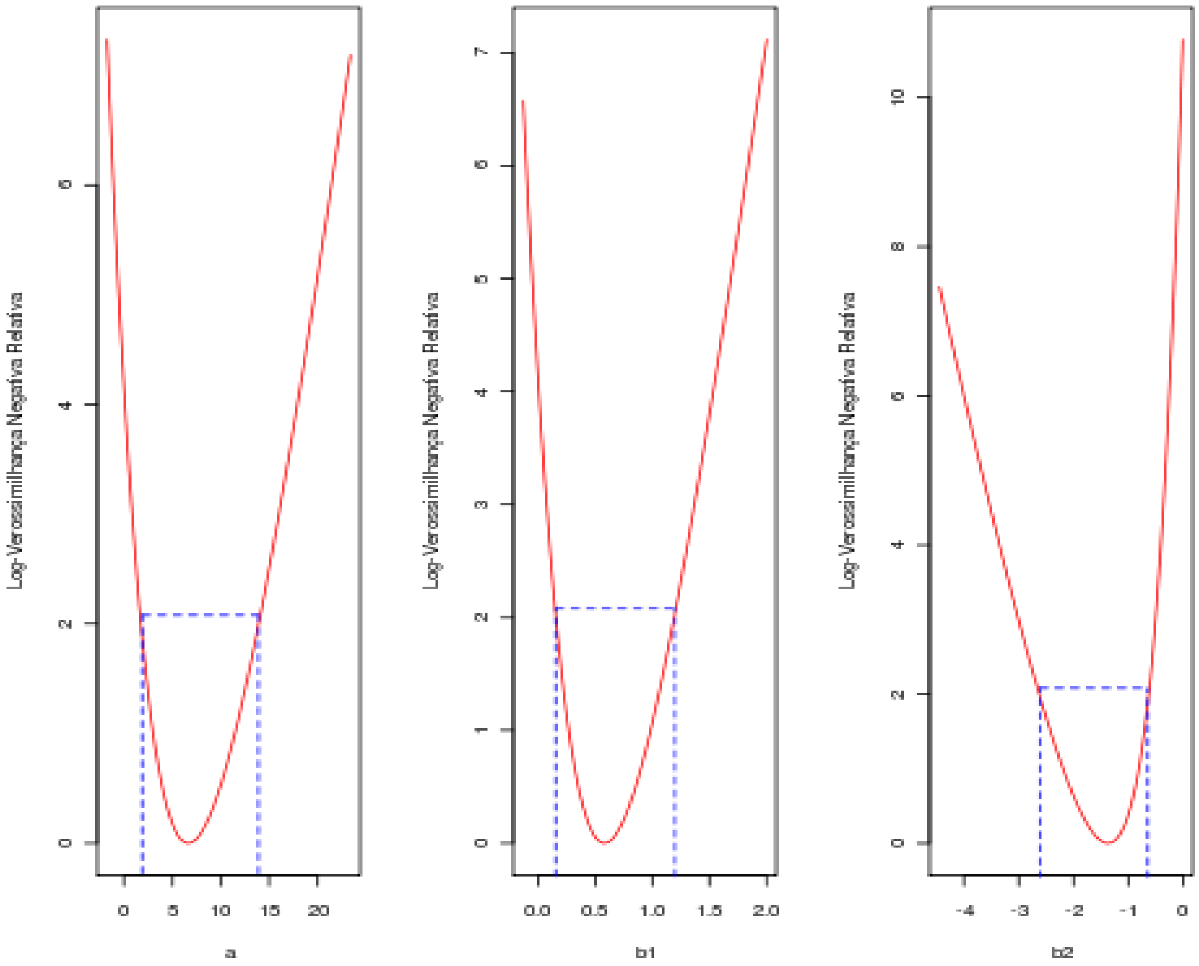






Modelos lineares generalizados de distribuição binomial

- Perfis de verossimilhança, Intervalos de verossimilhança.



Conclusões: Crawley

```
Coefficients:
              Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)     6.6417      2.9218    2.273   0.02302  *
area            0.5807      0.2478    2.344   0.01909  *
isolation      -1.3719      0.4769   -2.877   0.00401  **
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68.029 on 49 degrees of freedom
Residual deviance: 28.402 on 47 degrees of freedom
```

The estimates and their standard errors are in logits. Area has a significant positive effect (larger islands are more likely to be occupied), but isolation has a very strong negative effect (isolated islands are much less likely to be occupied). This is the minimal adequate model. We should plot the fitted model through the scatterplot of the data. It is much easier to do this for each variable separately, like this:

Modelos lineares generalizados de distribuição binomial, ligação probit

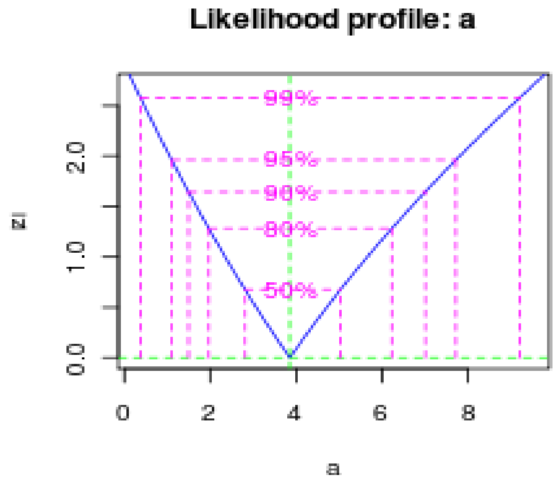
- Obtendo a máxima verosimilhança e optimizando usando uma função de probabilidade acumulada normal

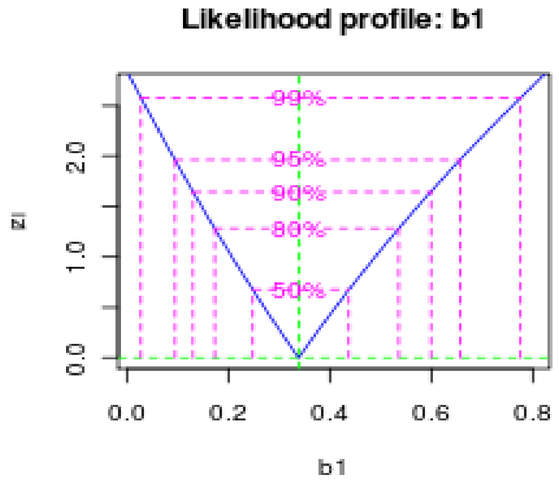
$$\Phi^{-1}\left[\pi(x)\right] = \alpha + \beta x$$

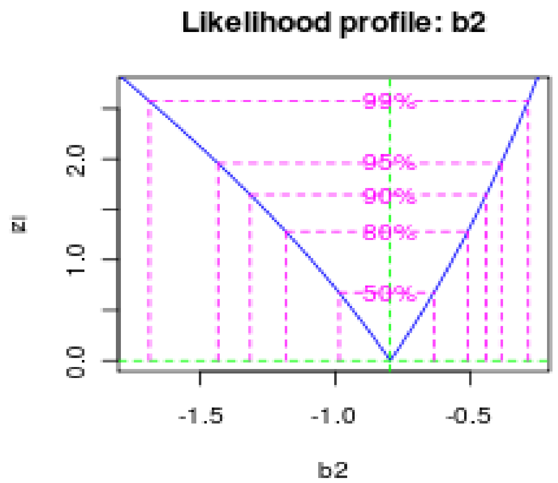
- `chalmod.pb1<-mle2(incidence~dbinom(pnorm(a + b1*area + b2*isolation), size=1), start=list(a=1, b1=0.05, b2 = -0.13))`

Modelos lineares generalizados de distribuição binomial,ligação probit

- Perfis de verossimilhança, premissas I. confiança.

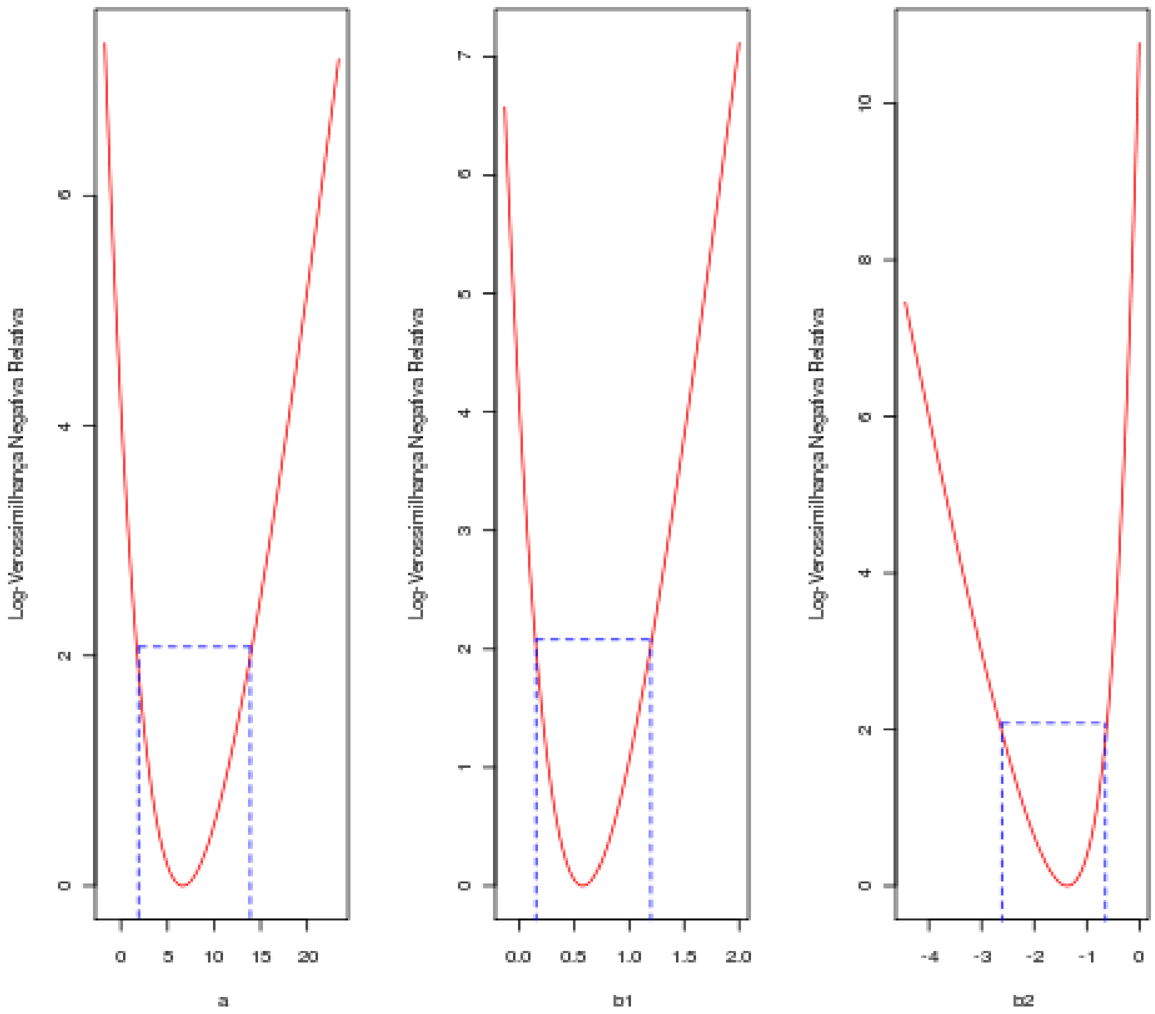






Modelos lineares generalizados de distribuição binomial,ligação probit

- Perfis de verossimilhança, premissas l. verosimilhança.



Modelos lineares generalizados de distribuição binomial,ligação clog-log

$$\pi(x) = 1 - \exp[-\exp(\alpha + \beta x)]$$

$$\log[-\log(1 - \pi(x))] = \alpha + \beta x.$$

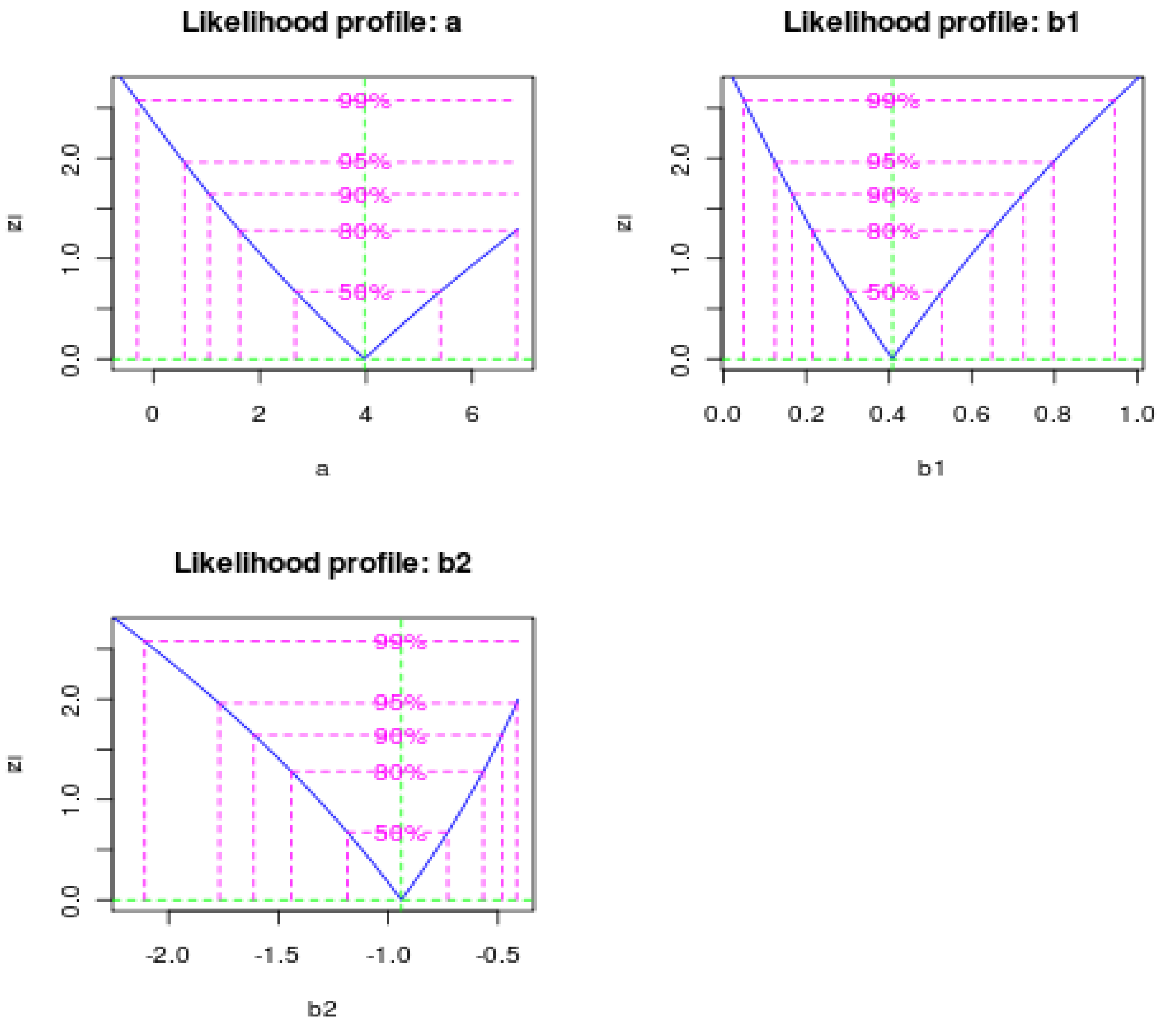
- `complog<-function(x) (1-exp(-exp(x)))`

Modelos lineares generalizados de distribuição binomial,ligação clog-log

- Optimizando
- `chalmod.clog1<-mle2(incidence~dbinom(complog(a + b1*area + b2*isolation), size=1), start=list(a=3, b1=0.4, b2 = -0.8))`

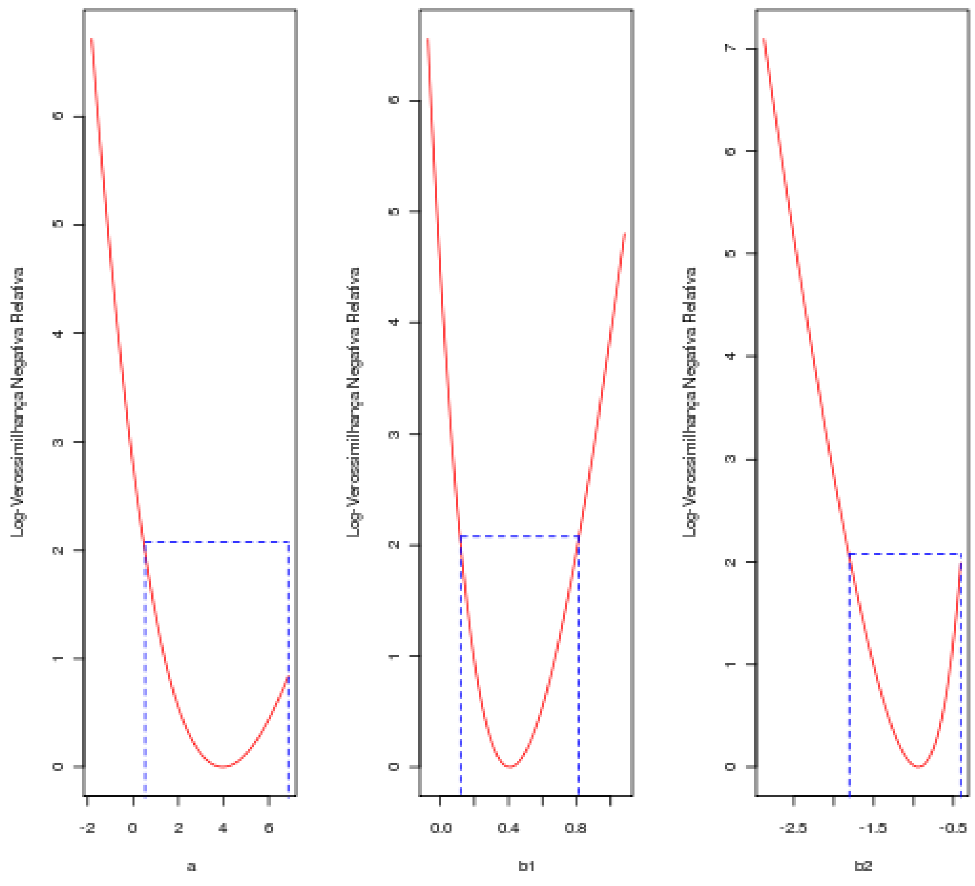
Modelos lineares generalizados de distribuição binomial,ligação clog-log

- Perfil,Intervalo de confiança



Modelos lineares generalizados de distribuição binomial, ligação clog-log

- Perfil,Intervalo de verossimilhança



Modelos lineares generalizados de distribuição binomial

- Seleção de modelos

- `AICctab(chalmod.lg2,chalmod.pb1, chalmod.clog1,weights=TRUE, nobs=50)`

- Modelos com suporte similar.

```
              AICc df weight
```

- `chalmod.lg2   34.9 3  0.290`
- `chalmod.pb1   34.5 3  0.357`
- `chalmod.clog1 34.5 3  0.353`

Modelos lineares generalizados de distribuição binomial

- Seleção do modelos

- `AICctab(chalmod.lg2,chalmod.pb1,weights=TRUE, nobs=50)`

- Modelos com suporte similar.

	AICc	gl	peso
Distancia + área	34.5	3	0.7202
Distancia * área	36.7	4	0.2404
Distancia	40.3	2	0.0394
Área	54.0	2	<0.001
Nulo	70.1	1	<0.001

Modelos lineares generalizados de distribuição binomial

- Superfície de resposta probit