



RODOLFO DE SOUZA NASCIMENTO

**Avaliação da maturação de peixes com  
regressões binárias: explorando novas funções  
de ligações**

Maringá – PR

2024

RODOLFO DE SOUZA NASCIMENTO

## **Avaliação da maturação de peixes com regressões binárias: explorando novas funções de ligações**

Dissertação apresentada ao Programa de Pós-graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

**Orientador:** Prof. Dr. Diego Corrêa Alves, Universidade Estadual de Maringá

**Membro:** Prof. Dr. Luis Carlos Gomes, Universidade Estadual de Maringá

**Membro:** Prof. Dr. Willian Luís de Oliveira, Universidade Estadual de Maringá

Universidade Estadual de Maringá - UEM

Departamento de Estatística - DES

Programa de Pós-Graduação em Bioestatística

Maringá-PR

2024

RODOLFO DE SOUZA NASCIMENTO

**Avaliação da maturação de peixes com regressões binárias: explorando novas funções de ligações**

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

**BANCA EXAMINADORA**

---

Prof. Dr. Diego Corrêa Alves

Universidade Estadual de Maringá – UEM

---

Prof. Dr. Luis Carlos Gomes

Universidade Estadual de Maringá – UEM

---

Prof. Dr. Willian Luís de Oliveira

Universidade Estadual de Maringá – UEM

Maringá, 28 de março de 2025

## AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão aos meus pais, Ademir Melo e Elenice Souza, e aos meus irmãos, Rodrigo Souza e Victória Souza, pelo constante incentivo aos estudos, pelo apoio incondicional ao longo desses anos de mestrado e por terem sido fundamentais na minha formação como ser humano.

Agradeço a Hagata Nathielly por tudo que fez para me ajudar nesta jornada que sabemos que não foi fácil, e sua família, Ricardo Almancio e Giseli Moreira pelo suporte pessoal e por serem uma família que nunca mediram esforços para me ajudar no que precisasse. Estendo este agradecimento em especial para Vó Neusa, sou seu fã por “n” motivos.

Agradeço a todos os conhecidos e desconhecidos no Amapá que me ajudaram na vaquinha que fiz para poder vir para Maringá-PR realizar este sonho de me tornar mestre e aos meus amigos, Carla Pacheco, Vinicius Neri, Ailton Costa, Clinton Agnes, Victor Keven, Matheus Campos, Felipe Batista, Seme Campos, Thiago Campos, Taissa Batista, Ana Moreira e muitos outros(as), pelo suporte pessoal e pela amizade durante esses anos.

Aos professores e colegas de trabalho, Professor Ricardo Puziol, Professor Paulo, por sempre apoiarem minhas decisões, incentivarem minha carreira acadêmica e por serem amigos nesta trajetória.

Um agradecimento especial ao meu orientador, Professor Diego Corrêa Alves, pelas conversas descontraídas, pela liberdade concedida na condução desta dissertação, pelo suporte no desenvolvimento deste trabalho, pela paciência, pela orientação no campo da Estatística/Computação e, acima de tudo, por acreditar no meu sucesso em todos os momentos.

Minha gratidão também à coordenação e ao corpo docente do Programa de Pós-Graduação em Bioestatística, pelo conhecimento transmitido, pela dedicação e excelência que fazem deste programa um destaque entre os demais.

Agradeço ao Núcleo de Pesquisas em Limnologia, Ictiologia e Aquicultura (NUPELIA) e a todos os seus funcionários, por disponibilizarem os dados essenciais para a realização deste trabalho.

## RESUMO

A maturação dos peixes é um fator essencial para o manejo pesqueiro e a definição de tamanhos mínimos de captura. Este estudo apresenta novas abordagens estatísticas para o estudo de maturação de peixes. São propostas diferentes funções de ligação para regressões binárias, incluindo as funções potência e reversa potência, sob a perspectiva bayesiana. Embora as funções de ligação logit e probit sejam amplamente utilizadas, elas podem ser inadequadas caso não capturem adequadamente a estrutura dos dados, resultando em vieses nas estimativas. Assim, funções potência e reversa potência, derivadas da generalização de distribuições normais, foram investigadas como alternativas mais flexíveis, especialmente para dados desbalanceados. Neste estudo, os modelos propostos foram ajustados aos dados de uma população da espécie *Prochilodus lineatus*, proveniente da planície de inundação do Alto Rio Paraná. Assim, foram estimados os parâmetros L50 e L99 com base em cada modelo, com diferentes funções de ligação. A comparação dos modelos foi realizada com base no *Deviance Information Criterion* (DIC), análise gráfica, análise de resíduos via envelope simulado e no tempo de processamento. Os resultados indicam que a função de ligação *potência cauchit* (PC) apresentou o melhor ajuste segundo o DIC e o envelope simulado, mas estimou inadequadamente o L99. Dessa forma, a função *reversa de potência logit* (RPL) foi identificada como uma opção mais equilibrada, combinando bom ajuste para L50 e L99, desempenho adequado no envelope simulado e o segundo melhor DIC.

**Palavras-chave:** Curimba, Sábalo, Planície de inundação, regressão logística, pesca.

## ABSTRACT

Fish maturation is an essential factor for fisheries management and the determination of minimum capture sizes. This study presents new statistical approaches for studying fish maturation. Different link functions for binary regressions are proposed, including power and reverse power functions, from a Bayesian perspective. Although logit and probit link functions are widely used, they may be inadequate if they do not properly capture the data structure, leading to biased estimates. Thus, power and reverse power functions, derived from the generalization of normal distributions, were investigated as more flexible alternatives, especially for unbalanced data. In this study, the proposed models were fitted to data from a population of the species *Prochilodus lineatus*, originating from the Upper Paraná River floodplain. The L50 and L99 parameters were estimated based on each model, using different link functions. Model comparisons were performed based on the Deviance Information Criterion (DIC), graphical analysis, residual analysis via simulated envelope, and processing time. The results indicate that the power cauchit (PC) link function provided the best fit according to the DIC and simulated envelope but inadequately estimated L99. Therefore, the reverse power logit (RPL) function was identified as a more balanced option, combining a good fit for L50 and L99, adequate performance in the simulated envelope, and the second-best DIC.

**Keywords:** Curimba, Sábalo, Floodplain, logistic regression, fishing.

## LISTA DE FIGURAS

<b>Figura 1:</b> Mapa da planície do Rio Paraná.....	19
<b>Figura 2:</b> Relação entre o comprimento padrão e a proporção de indivíduos maturados .....	26
<b>Figura 2:</b> Curvas ajustadas da probabilidade de maturação para a fêmea em função do comprimento padrão .....	31
<b>Figura 3:</b> Curvas ajustadas da probabilidade de maturação para o macho em função do comprimento padrão .....	31
Relação entre o comprimento padrão e a proporção de indivíduos maturados	

## LISTA DE TABELAS

<b>Tabela 1:</b> Expressões algébricas para L50 e L99 das funções de ligações .....	22
<b>Tabela 2:</b> Distribuição dos indivíduos por sexo e estágio de maturação, com proporções Relativas à maturação .....	26
<b>Tabela 3:</b> Valores do Critério de Informação da <i>Deviance Bayesiano (DIC)</i> , medidas preditivas e tempo de processamento (TP).....	27
<b>Tabela 4:</b> Valores do comprimento médio para primeira maturação L50, L99, suas respectivas diferenças com intervalos HPD .....	29

## SUMÁRIO

1. INTRODUÇÃO.....	17
2. MATERIAL E MÉTODOS .....	19
2.1. Área de estudo.....	19
2.2. Amostragem de dados.....	19
2.3. Análise de dados.....	20
3. RESULTADOS .....	25
4. DISCUSSÃO.....	32
5. CONSIDERAÇÕES FINAIS.....	34
REFERÊNCIAS .....	36
APÊNDICES.....	42

## 1. INTRODUÇÃO

A maturação sexual dos peixes é um processo fundamental para a compreensão da dinâmica populacional e para a gestão sustentável dos estoques pesqueiros. Um peixe é considerado maturo quando suas gônadas atingem um estágio de desenvolvimento no qual a reprodução é biologicamente possível (Brown-Peterson *et al.*, 2011), estes descrevem um ciclo de maturação com várias fases: imaturo, em maturação, maturo, desovando e pós-desova (VAZZOLER, 1996). O estudo da maturação de peixes visa estimar parâmetros populacionais referentes ao tamanho em que os peixes estão aptos a se reproduzirem, sendo importantes para embasar práticas de manejo pesqueiro, como a definição de tamanhos mínimos de captura e a proteção de períodos de reprodução Agostinho *et al.* (2004).

A determinação do comprimento no qual 50% da população atinge a maturidade sexual (L50) é essencial para definir medidas de conservação e manejo Silva *et al.* (2021). O termo (L50) refere-se ao comprimento corporal de um peixe no qual metade dos indivíduos de uma população atingiu a maturidade sexual. É uma medida amplamente utilizada em estudos de biologia pesqueira para definir tamanho mínimo de captura (KING, 2007).

Além do L50, o L99 é o comprimento corporal no qual 99% da população de peixes atinge a maturidade sexual. Esse valor também é utilizado como um indicador para estudos de maturação de peixes, reflete o tamanho em que quase todos os indivíduos estão prontos para reproduzir. Na obra de King (2007) a definição de L99 é frequentemente empregada em estudos de biologia pesqueira para assegurar que as práticas de captura não interfiram nos ciclos reprodutivos das espécies.

Para estimar esse parâmetro, modelos de regressão binária são amplamente utilizados, pois permitem modelar a probabilidade de um indivíduo estar maturo em função de variáveis explanatórias, como o comprimento padrão e o sexo. Tradicionalmente, a função *logit* é utilizada por sua interpretação intuitiva e eficiência computacional (BRESLOW & CLAYTON, 1993; DEMÉTRIO, 2001; NUNES *et al.*, 2004; VAZZOLER, 1996).

No entanto, dependendo das características dos dados, outras funções de ligação podem oferecer melhor ajuste, permitindo maior flexibilidade na modelagem da maturação dos peixes.

De maneira geral, os estudos de maturação de peixes são realizados com as seguintes abordagens: análise macroscópica das gônadas, análise histológica, índice gonadosomático e estimativa do comprimento de primeira maturação sexual (L50). Entre essas, tem se destacado a utilização de curvas logísticas, como discutido por

Morgan (2008), que analisou diferentes métodos para estimar parâmetros de maturação e sua aplicação na gestão pesqueira, a variável resposta em estudos de maturação de peixes geralmente é binária, representando o estado de maturidade sexual do peixe: "1" se o peixe é maturo e "0" se é imaturo. Esta forma de abordar o problema acaba culminando em uma regressão binária, dentro da Teoria dos Modelos Lineares Generalizados (Nelder e Wedderburn, 1992), onde se considera a função *logit* como função de ligação canônica, resultando na regressão logística.

A regressão logística é uma técnica que também é utilizada em estudos de maturação, e a interpretação de seus parâmetros é realizada através da razão de chances (*Odds Ratio* - OR) (CHEN; DEY; SHAO, 1999; MCCULLAGH; NELDER, 1989).

A função *logit* é simétrica, implicando que a probabilidade de sucesso da variável resposta aumenta à mesma taxa que a probabilidade de fracasso diminui. Outras funções de ligação simétricas, como *probit* e *cauchit*, derivadas das distribuições Normal e Cauchy, também podem ser utilizadas (ALBERT; CHIB, 1993).

No entanto, o uso inadequado da função *logit* pode gerar viés nas estimativas dos parâmetros e nas probabilidades preditas, especialmente quando a função de ligação não reflete corretamente a natureza dos dados (CZADO; SANTNER, 1992).

Para lidar com dados assimétricos, funções de ligação como *loglog* e *cloglog* são alternativas derivadas das distribuições Gumbel de Valor Extremo Máximo e Mínimo, respectivamente, permitindo modelar assimetrias, embora com limitações de flexibilidade. A necessidade de funções de ligação mais flexíveis motivou o desenvolvimento de novas abordagens, isto é, à sua capacidade de se adaptar melhor a diferentes padrões de dados, como os dados desbalanceados, especialmente quando os dados apresentam comportamentos ou características que uma função de ligação mais simples (como a *logit* ou *probit*) pode não capturar adequadamente. As funções de potência e reversa potência, baseadas na generalização da distribuição normal proposta por Gupta e Gupta (2008), foram exploradas por Abanto-Valle, Bazán e Smith (2014), Bazán, Romeo e Rodrigues (2014) e Bazán *et al.* (2017).

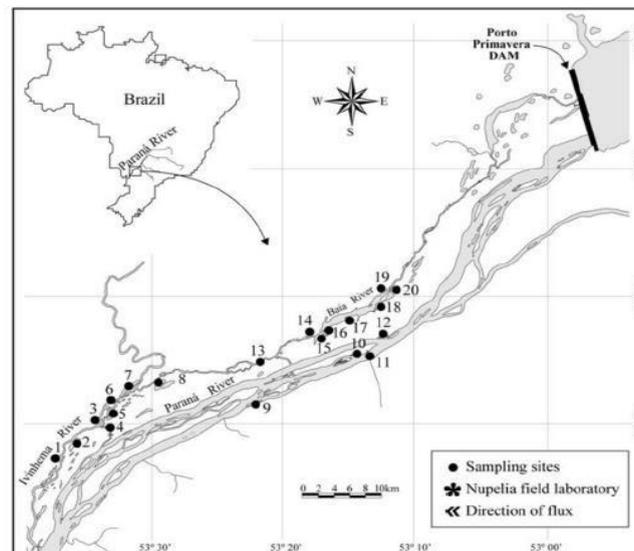
Portanto, este trabalho propõe explorar outras funções de ligação além da tradicional (*logit*): *probit*, *cauchit*, *potência logit*, *potência probit*, *potência cauchit*, *potência cloglog*, *potência loglog*, *reversa de potência logit*, *reversa de potência probit*, *reversa de potência cauchit*, *reversa de potência cloglog* e *reversa de potência loglog*.

Este estudo tem como objetivo avaliar o desempenho de modelos contendo diferentes funções de ligação na modelagem da maturação sexual de peixes, utilizando como estudo de caso a população de curimba (*Prochilodus lineatus*) da planície de inundação do Alto rio Paraná. A identificação do modelo com a função de ligação mais adequada contribuirá para uma melhor compreensão da dinâmica reprodutiva da espécie, fornecendo informações que poderão ser aplicadas a outros estudos de maturação e manejo pesqueiro.

## 2. MATERIAL E MÉTODOS

### 2.1. Área de estudo

O Rio Paraná é o principal rio da Bacia do Prata e o segundo mais longo da América do Sul. A planície de inundação do Alto Rio Paraná está localizada na margem oeste do Rio Paraná entre os estados de Mato Grosso do Sul e Paraná, Brasil, em um remanescente de 230 km de extensão com características lólicas, ou seja, ambientes de água corrente, onde está água fica em movimento constante entre a represa de Porto Primavera (a montante) e o reservatório de Itaipu (a jusante), (Souza Filho & Stevaux, 2004).



**Figura 1:** Mapa da planície do Rio Paraná.

Fonte: Vasconcelos, *et al.* (2014).

Na figura 1 encontra-se o rio Paraná, nesta é possível identificar os pontos de coleta a respeito do banco de dados analisado.

### 2.2. Amostragem de dados

Neste estudo, utiliza-se um conjunto de dados composto por 3 variáveis e 15.715 observações a respeito das características do peixe de espécie *Prochilodus lineatus* provenientes do Núcleo de Pesquisas em Limnologia, Ictiologia e Aquicultura (NUPELIA). Esses dados incluíam informações essenciais sobre a espécie, como:

comprimento padrão, sexo e estágio reprodutivo.

O Núcleo de Pesquisa em Limnologia, Ictiologia e Aquicultura (NUPELIA) coletou os dados por meio de amostras retiradas dos canais secundários, lagos conectados e não conectados e áreas lóxicas dos três principais rios do Alto Paraná. As amostragens de peixes foram realizadas com redes de emalhar que foram fixadas por 24 horas e verificadas a cada 8 horas. Todos os peixes capturados foram identificados de acordo com Graça & Pavanelli (2007).

A partir dos dados obtidos, que tem como base os dados de maturação do peixe de espécie *Prochilodus lineatus*, foram analisadas variáveis relacionadas à sua maturação, ao sexo, e comprimento padrão para verificar quais estiveram associadas à ocorrência da maturação do peixe. Um total de 15.715 registros de maturação da espécie coletada nos três principais rios do Alto Paraná, entre os anos 1986 e 2018 e informações de duas possíveis variáveis explicativas (fatores) foram utilizadas.

### 2.3. Análise dos dados

A maturação dos peixes foi modelada por meio de modelos lineares generalizados (Nelder e Wedderburn, 1992). A variável resposta, a maturação ( $Y$ ), foi obtida definindo-se como “maturo”,  $Y = 1$ , o peixe que se encontra nesta fase no momento de sua identificação,  $Y = 0$ , caso contrário. Assim, considere  $Y = (y_1, y_2, \dots, y_n)^T$ , o vetor de variáveis aleatórias independentes que indicam a maturação dos peixes da espécie *Prochilodus lineatus*, de forma que  $Y \sim Ber(\theta)$ , no qual  $P(Y = 1) = \theta$  representa a probabilidade de um peixe estar maturo.

Seja a matriz  $X = [Comprimento_i (cm), Sexo_i]$ , com  $i = 1, \dots, n$ . representando o conjunto de vetores dos valores correspondentes às duas variáveis explanatórias: comprimento padrão e sexo. Isso forma a matriz de incidência  $X = [1, x_1, x_2]$  do modelo. Assim, um modelo de regressão binária é especificado da seguinte forma:

$$Y | x \stackrel{iid}{\sim} Ber(\theta)$$

$$\theta = E[Y|x] = F(\eta) = F(x\beta)$$

ou equivalentemente:

$$\eta = x\beta = F^{-1}(\theta)$$

onde  $\beta = (\beta_0, \dots, \beta_k)^T$  é o vetor de coeficientes de regressão,  $F(\cdot)$  é a função distribuição acumulada (f.d.a.) e  $F^{-1}(\cdot)$  a função de ligação (também comumente expressa como  $g(\cdot)$ ) que relaciona  $\theta$  à  $\eta$ , o preditor linear, o qual determina a combinação linear entre as variáveis dependentes e os coeficientes da regressão, dada por:  $\eta_i = \beta_0 + \beta_1 \times Comprimento_i + \beta_2 \times Sexo_i$ , sendo  $i = 1, \dots, 15.715$  indivíduos e  $k = 2$  variáveis explanatórias: comprimento padrão e sexo. Nesse caso,  $\beta_0$  representa o

logaritmo das chances (log-odds) de um indivíduo estar maturo quando o comprimento é igual a zero. Biologicamente, isso corresponderia a peixes em fase larval, nos quais a probabilidade de maturação é praticamente nula, independentemente do sexo. Assim, incluir a variável sexo de forma isolada  $\beta_3 \times Sexo_i$  não traria ganho interpretativo, já que não há diferença de maturação entre machos e fêmeas em comprimentos tão pequenos. Além disso, a inclusão do termo  $\beta_3 \times Sexo_i$  aumentaria a correlação entre os parâmetros, complicando o modelo sem agregar valor científico. Portanto, optou-se por incluir apenas o termo de interação entre comprimento e sexo, que reflete adequadamente diferenças de maturação entre os sexos em comprimentos relevantes.

Com relação às funções de ligação, foram utilizadas as cinco funções básicas, logit, probit, Cauchy, cloglog e loglog, sendo as três primeiras de característica simétrica e obtidas das distribuições de probabilidade Logística, Normal e Cauchy, respectivamente, e as duas últimas, de característica assimétrica, obtidas das distribuições Gumbel de Valor Mínimo e Máximo, respectivamente (Aranda-ordaz, 1981). Todas estas funções podem ser observadas na tabela A1, do material suplementar (Apêndice A). Além dessas funções de ligação usuais, também foram utilizadas as classes de funções de ligação conhecidas como potência e reversa de potência, apresentadas na tabela A2 do material suplementar (Apêndice A). Além disso, o Apêndice A apresenta uma fundamentação teórica sobre o assunto de regressão binária e as funções de ligação utilizadas.

Para determinar as medidas de comprimento sexual L50 e L99 foram utilizadas as funções de ligação descritas no material suplementar (Apêndice A), as expressões destas medidas estão abaixo (Tabela 1).

**Tabela 1:** Expressões algébricas para L50 e L99 das funções de ligações.

Origem	Nome	<sup>1</sup> L50	<sup>2</sup> L99
Logística	<b>Logit (L)</b>	$-\frac{\beta_0}{\beta_1}$	$\frac{\log(99) - \beta_0}{\beta_1}$
Normal	<b>Probit (P)</b>	$-\frac{\beta_0}{\beta_1}$	$\frac{\phi^{-1}(0,99) - \beta_0}{\beta_1}$
Cauchy	<b>Cauchit (C)</b>	$-\frac{\beta_0}{\beta_1}$	$\frac{\tan(\pi(0,49)) - \beta_0}{\beta_1}$
Gumbel de Valor mínimo	<b>Cloglog (CLL)</b>	$\frac{\log(-\log(0,5)) - \beta_0}{\beta_1}$	$\frac{\log(-\log(0,01)) - \beta_0}{\beta_1}$
Gumbel de valor máximo	<b>Loglog (LL)</b>	$\frac{-\log(-\log(0,5)) - \beta_0}{\beta_1}$	$\frac{-\log(-\log 0,99) - \beta_0}{\beta_1}$
Potência Logit (PL)	<b>PL</b>	$\frac{\log\left(\frac{0,5^{\frac{1}{\lambda}}}{1 - 0,5^{\frac{1}{\lambda}}}\right) - \beta_0}{\beta_1}$	$\frac{\log\left(\frac{0,99^{\frac{1}{\lambda}}}{1 - 0,99^{\frac{1}{\lambda}}}\right) - \beta_0}{\beta_1}$
Potência Probit (PP)	<b>PN</b>	$\frac{\phi^{-1}\left(0,5^{\frac{1}{\lambda}}\right) - \beta_0}{\beta_1}$	$\frac{\phi^{-1}\left(0,99^{\frac{1}{\lambda}}\right) - \beta_0}{\beta_1}$
Potência Cauchit (PC)	<b>PC</b>	$\frac{\tan\left(\pi\left(0,5^{\frac{1}{\lambda}} - 0,5\right)\right) - \beta_0}{\beta_1}$	$\frac{\tan\left(\pi\left(0,99^{\frac{1}{\lambda}} - 0,99\right)\right) - \beta_0}{\beta_1}$
Potência Cloglog (PCLL)	<b>PGVMi</b>	$\frac{\log\left(-\log\left(1 - 0,5^{\frac{1}{\lambda}}\right)\right) - \beta_0}{\beta_1}$	$\frac{\log\left(-\log\left(1 - 0,99^{\frac{1}{\lambda}}\right)\right) - \beta_0}{\beta_1}$
Potência loglog (PLL)	<b>PGVMa</b>	$\frac{-\log\left(-\log\left(0,5^{\frac{1}{\lambda}}\right)\right) - \beta_0}{\beta_1}$	$\frac{-\log\left(-\log\left(0,99^{\frac{1}{\lambda}}\right)\right) - \beta_0}{\beta_1}$
Reversa de Potência logit (RPL)	<b>RPL</b>	$\frac{\log\left(\frac{1 - (0,5)^{\frac{1}{\lambda}}}{(0,5)^{\frac{1}{\lambda}}}\right) - \beta_0}{\beta_1}$	$\frac{\log\left(\frac{1 - (0,99)^{\frac{1}{\lambda}}}{(0,99)^{\frac{1}{\lambda}}}\right) - \beta_0}{\beta_1}$
Reversa de Potência Probit (RPP)	<b>RPP</b>	$\frac{-\phi^{-1}\left((0,5)^{\frac{1}{\lambda}}\right) - \beta_0}{\beta_1}$	$\frac{-\phi^{-1}\left((0,99)^{\frac{1}{\lambda}}\right) - \beta_0}{\beta_1}$
Reversa de Potência Cauchit (RPC)	<b>RPC</b>	$\frac{-\tan\left(\pi\left((0,5)^{\frac{1}{\lambda}} - 0,5\right)\right) - \beta_0}{\beta_1}$	$\frac{-\tan\left(\pi\left((0,99)^{\frac{1}{\lambda}} - 0,99\right)\right) - \beta_0}{\beta_1}$
Reversa de Potência Cloglog (RPCLL)	<b>RPCLL</b>	$\frac{-\log\left(-\log\left(1 - (0,99)^{\frac{1}{\lambda}}\right)\right) - \beta_0}{\beta_1}$	$\frac{-\log\left(-\log\left(1 - (0,99)^{\frac{1}{\lambda}}\right)\right) - \beta_0}{\beta_1}$
Reversa de Potência Loglog (RPLL)	<b>RPLL</b>	$\frac{\log\left(-\log\left(0,5^{\frac{1}{\lambda}}\right)\right) - \beta_0}{\beta_1}$	$\frac{\log\left(-\log\left(0,99^{\frac{1}{\lambda}}\right)\right) - \beta_0}{\beta_1}$

<sup>1</sup>L50: Comprimento corporal de um peixe no qual metade dos indivíduos de uma população atingiu a maturidade sexual.

<sup>2</sup>L99: Comprimento corporal no qual 99% da população de peixes atinge a maturidade sexual.

Os modelos descritos foram ajustados sob a perspectiva Bayesiana. Segundo DA SILVA (2020), nos modelos que utilizam funções de ligação básicas, os parâmetros a serem estimados correspondem aos coeficientes de regressão  $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ . Já para as funções de ligação potência e reversa de potência, além desses coeficientes, é

necessário estimar o parâmetro de assimetria  $\lambda$ , cuja estimação é facilitada pela transformação  $\delta = \ln(\lambda)$ , equivalente a  $\lambda = \exp(\delta)$ . Essa transformação, conforme Bazán et al. (2014), simplifica os cálculos em abordagens bayesianas.

As distribuições *a priori* foram definidas de forma a carregar pouca informação (priors vagas), como segue:

$$Y | \beta, \alpha \stackrel{iid}{\sim} Ber(\theta),$$

$$\theta = F(\eta);$$

$$\beta \sim N(\mu = 0, \tau = 10^{-6}) \text{ e } \delta \sim U(-2, 2) \rightarrow \lambda(0,14; 7,39).$$

Note que os modelos com funções de ligação potência e reversa potência são reduzidos para os modelos com funções de ligação básicas quando  $\delta = 0$  e  $\lambda = 1$  (para mais detalhes, consultar o Apêndice A). Devido à parametrização Jags parallel (Plummer, 2006) (programa computacional para inferência em modelos bayesianos), a distribuição *a priori* para  $\beta$  foi uma Normal com hiperparâmetros média  $\mu = 0$  e precisão  $\tau = 10^{-6}$  (com variância  $\sigma^2 = \tau^{-1}$ ). A distribuição *a priori* para  $\delta$ , foi definida como uma Uniforme com hiperparâmetros  $a = -2$  e  $b = 2$ , já que os valores de  $\lambda$  no intervalo  $[e^{-2}, e^2]$  apresentam uma maior probabilidade de ocorrência para dados desbalanceados (Da Silva, 2020).

As distribuições *a posteriori* dos parâmetros foram obtidas numericamente por meio do software JAGS (*Just Another Gibbs Sampler*) (Plummer, 2012), que é chamado indiretamente pelo R, utilizando o ambiente RStudio (R Core Team, 2024) por meio do pacote rjags (Plummer, 2023). Para melhorar a eficiência computacional, a execução das cadeias foi paralelizada utilizando o pacote JAGSparallel (Plummer, 2023). Para os modelos clássicos utilizamos como valores iniciais para  $\beta$  estimativas frequentistas com suas respectivas ligações e  $\lambda = 0$ . E para os modelos de potência e reversa de potência utilizamos para  $\lambda$ , um valor gerado aleatoriamente de uma distribuição Uniforme, com parâmetros  $a = -2$  e  $b = 2$ . As distribuições *a posteriori* foram obtidas por meio da técnica de Monte Carlo Cadeias de Markov (MCMC: *Markov Chain Monte Carlo*, do inglês). Neste processo, foram gerados 300.000 valores, considerando um período de descarte inicial (burn-in) de 100.000 e saltos de tamanho 400 para reduzir a autocorrelação gerada no processo MCMC, utilizando 6 cadeias. Assim, após este processo, obteve-se amostras finais de 3.000 valores gerados das distribuições *a posteriori*. A convergência das cadeias foi verificada pelos critérios de Geweke (Geweke, 1992) e Heidelberger e Welch (Heidelberger & Welch, 1983), implementados na biblioteca R “coda” (PLUMMER, 2006). Adicionalmente, a convergência foi avaliada

graficamente por meio das funções *traplot* e *denplot*, que permitiram inspecionar visualmente a estabilidade das cadeias e a forma da densidade posterior. A autocorrelação foi avaliada por meio da função *acf* (*autocorrelation function*), também disponível no pacote “coda”.

As estimativas dos parâmetros, obtidas a partir das distribuições *a posteriori*, foram as medidas pontuais: média, desvio-padrão e mediana, também foram calculados L50 (comprimento de 50%) e L99 (comprimento de 99%, como *proxy* do L100),  $\lambda$ ,  $\delta$ , Diff's (diferença entre os comprimentos entre machos e fêmeas para o L50 e L99) e seus respectivos intervalos HPD, todas estas estimativas podem ser observadas nas tabelas B1, B2 e B3 do material suplementar (Apêndice B). As distribuições para os tamanhos de maturação (L50 e L99) foram obtidas para cada modelo (função de ligação), como segue na tabela 1. A significância dos parâmetros estimados foi verificada caso os seus, respectivos, intervalos HPD (*Highest Posterior Density*), com 95% de credibilidade, não contivessem o valor zero.

Para a seleção dos modelos finais, foi utilizado o critério de informação baseado na *Deviance* (DIC: *Deviance Information Criterion*) (SPIEGELHALTER, 2002). Esse critério permite avaliar a parcimônia do modelo, ou seja, a combinação entre a qualidade do ajuste e sua complexidade. Quanto menor o valor de DIC, mais parcimonioso o modelo será. A diferença entre os DICs ( $\Delta$ DIC) de dois modelos quantifica o quão superior é o ajuste de um modelo em relação ao outro, sendo que o modelo com  $\Delta$ DIC igual a zero é considerado o mais parcimonioso e, portanto, o melhor candidato a modelo final. Valores positivos de  $\Delta$  DIC indicam que o modelo com menor DIC deve ser escolhido. De acordo com Spiegelhalter et al. (2002), se a diferença entre os valores de DIC for inferior a cinco, os modelos podem ser considerados concorrentes, pois apresentam desempenho semelhantes.

Além do DIC, a avaliação da qualidade do ajuste dos modelos também pode ser feita por meio de medidas preditivas, que permitem comparar o desempenho dos diferentes ajustes. Entre essas medidas, destacam-se Sensibilidade (*s*), Especificidade (*e*) e Acurácia (ACC), que são utilizadas em testes diagnósticos e expressam a capacidade do modelo em classificar corretamente a ocorrência ou não do evento de interesse (GILOLO, 2017).

Para complementar esta análise, optou-se pelo uso do resíduo quantil aleatorizado, proposto por Dunn e Smyth (1996). Esse resíduo apresenta a propriedade de seguir uma distribuição normal padrão independentemente da distribuição da variável resposta, desde que o modelo esteja corretamente ajustado.

A análise gráfica dos resíduos pode ser realizada por meio de diferentes abordagens, Resíduos versus Índice da Observação e/ou Resíduos versus Valores Ajustados que servem para verificar valores discrepantes e a homogeneidade/média zero. De acordo com Paula (2013), o gráfico de envelope simulado tem se mostrado uma ferramenta para essa avaliação. Atkinson (1981) recomenda que as bandas de confiança do envelope sejam construídas por meio de simulações, pois a utilização direta da variância dos resíduos pode introduzir vieses na avaliação do modelo.

Estudos como os de Lemonte e Bazán (2018) e Huayanay (2019) já aplicaram resíduos quantis aleatorizados e gráficos de envelopes simulados para verificar a adequabilidade de modelos com funções de ligação potência e reversa de potência. Enquanto Lemonte e Bazán (2018) abordaram essa questão sob uma perspectiva frequentista, Huayanay (2019) utilizou uma abordagem bayesiana.

Além da análise dos resíduos, a eficiência computacional dos modelos também foi considerada, sendo mensurada pelo tempo de processamento (TP) necessário para a obtenção das distribuições a posteriori convergentes. Esse tempo, medido em horas, foi registrado para cada modelo, garantindo uma avaliação objetiva do custo computacional associado a cada ajuste. Para assegurar a reprodutibilidade dos resultados, cada modelo foi ajustado após a reinicialização do sistema operacional. Por fim, a adequabilidade dos resíduos foi avaliada por meio dos gráficos de índices e do envelope simulado, conforme recomendado por Atkinson (1981).

Todas as simulações, bem como a geração dos resultados gráficos e não gráficos, foram realizadas em uma máquina equipada com um processador 13th Gen Intel® Core™ Intel Core i5-1335U 1.30GHz e 16,0 GB de RAM. Este sistema operou sob um sistema operacional de 64 bits, utilizando o Windows 11 Pro.

### **3. RESULTADOS**

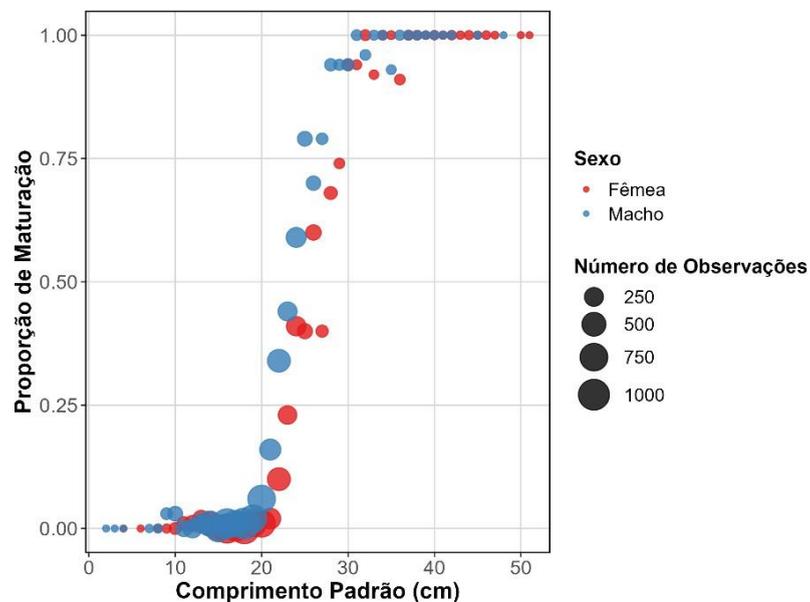
A Tabela 2 apresenta a distribuição dos indivíduos segundo o estágio de maturação (imaturo ou maturo) e o sexo. No total, foram analisados 15.230 indivíduos, sendo 7.262 fêmeas (47,68%) e 7.968 machos (52,32%). Observa-se um desbalanceamento na variável resposta, com a grande maioria dos indivíduos classificados como imaturos (88,62%) e apenas 11,38% como maturos.

**Tabela 2:** Distribuição dos indivíduos por sexo e estágio de maturação, com proporções relativas à maturação.

Sexo	Imaturo	Maturo	Total
Fêmea	6584 (90,66%)	678 (9,34%)	7262 (47,68%)
Macho	6913 (86,76%)	1055 (13,24%)	7968 (52,32%)
Total	13497 (88,62%)	1733 (11,38%)	15230

Dentre as fêmeas, 6.584 indivíduos (90,66%) foram classificados como imaturos, enquanto 678 (9,34%) apresentaram-se maturos. Já entre os machos, 6.913 indivíduos (86,76%) estavam no estágio imaturo e 1.055 (13,24%) no estágio maturo. Esse desbalanceamento entre as classes pode influenciar a modelagem, sendo um fator a ser considerado nas análises subsequentes.

O gráfico exposto em figura 2 apresenta a proporção de indivíduos maturados em relação ao comprimento padrão, diferenciados por sexo (fêmeas em vermelho e machos em azul), com o tamanho dos pontos representando o número de observações em cada classe de comprimento.



**Figura 2:** Relação entre o comprimento padrão e a proporção de indivíduos maturados de *prochilodus lineatus* coletados na planície de inundação do rio Paraná entre os anos de 1986 e 2018.

Observa-se uma grande quantidade de amostras nos menores comprimentos, especialmente abaixo de 20 cm, onde a proporção de indivíduos imaturos é predominante (próxima de 0).

Além disso, a densidade elevada de amostras nos comprimentos iniciais pode influenciar os modelos ajustados posteriormente, tornando as estimativas da curva de maturação mais sensíveis a essa concentração de dados. Isso pode impactar a precisão da L50 e L99 estimada e das comparações entre diferentes funções de ligação. No

intervalo aproximadamente entre 20 cm e 30 cm, observa-se uma maior dispersão na proporção de maturação, refletindo a variabilidade individual no desenvolvimento reprodutivo. Após esse ponto, a proporção se estabiliza próxima de 1, indicando que praticamente todos os indivíduos com comprimentos superiores a 30 cm estão maturados.

**Tabela 3:** Valores do Critério de Informação da *Deviance Bayesiano* (DIC), medidas preditivas e tempo de processamento (TP).

Tipo de ligação	Funções de ligação	DIC	$\Delta DIC$	Acurácia	Sensibilidade	Especificidade	TP
Básicas	L	1282,9	113	93,3	60,6	97,5	43 Min.
	P	1404,4	234,5	93	53,9	98,1	24 Min.
	C	1732,7	562,8	92,9	50,8	98,3	21 Min.
Potências	PL	1279,4	109,5	93	66,1	96,5	34 Min.
	PP	1404,3	234,4	93	53,9	98,1	49 Min.
	PC	<b>1169,9</b>	0	92,9	50,7	98,3	24 Min.
Reversas	RPL	1261,8	91,9	92,9	72,4	95,6	37 Min.
	RPP	1386,4	216,5	93	54,6	98	40 Min.
	RPC	1298,3	127,55	92,9	50,8	98,3	35 Min.

Todos os modelos ajustados passaram nos testes de Geweke e de Heidelberger e Welch, verificando, assim, a convergência nas cadeias *a posteriori*. A partir dos valores DIC e  $\Delta DIC$ , é possível destacar o modelo *Potência Cauchit* (PC) como mais parcimonioso, enquanto o segundo modelo mais parcimonioso foi o modelo *Reversa de Potência logit* (RPL), com  $\Delta DIC$  próximo a 92 (Tabela 3). Ainda, destaca-se que todos os modelos com ligações do tipo Potência e Reversa de Potência apresentaram menores valores de DIC do que a sua contraparte da família básica (Tabela 3), sugerindo que estes modelos parecem ter se ajustado melhor aos dados.

Com base no Critério de Informação da Deviance (DIC), observa-se que o ranking de menor para maior DIC segue a seguinte ordem: *potência cauchit* (PC), reversa de *potência logit* (RPL), *potência logit* (PL), *logit* (L) e reversa de *potência cauchit* (RPC).

No que se refere às medidas preditivas, os valores de acurácia foram próximos entre todos os modelos ajustados, indicando que, independentemente da função de ligação utilizada, os modelos apresentam um desempenho semelhante na classificação correta dos indivíduos imaturos e maturados. Entretanto, ao analisarmos as métricas separadamente, observa-se que a família *cauchit* obteve os maiores valores de sensibilidade, ou seja, maior capacidade de identificar corretamente os indivíduos maturados. Por outro lado, a família *probit* apresentou a maior especificidade, ou seja, maior capacidade de identificar corretamente os indivíduos imaturos. Esse resultado sugere que, caso o objetivo seja um modelo mais equilibrado entre sensibilidade e especificidade, a família *logit* pode ser a escolha mais apropriada. De acordo com o DIC, dentro dessa família, o modelo com a função de ligação *reversa de potência logit* (RPL)

foi o mais indicado.

A avaliação da qualidade do ajuste foi realizada por meio do envelope simulado, esta avaliação pode ser observada em figura D1, do material suplementar (Apêndice D). Observou-se que os modelos com melhor desempenho foram aqueles da família *cauchit*, seguidos pelos modelos *logit* e, por fim, pelos modelos *probit*.

Por fim, a eficiência computacional dos modelos também foi analisada. Os tempos de processamento variaram consideravelmente entre as diferentes funções de ligação, sendo que os modelos baseados na família *logit* apresentaram tempos intermediários, enquanto a família *cauchit* exigiu maior tempo computacional para convergência das distribuições a posteriori. Destaca-se, entre os modelos analisados, o tempo computacional específico dos modelos RPL e RPC, que apresentaram tempos intermediários, e mantiveram um equilíbrio entre qualidade de ajuste e parcimônia.

Analisando as medidas de primeira maturação L50 e L99 (Tabela 4) percebe-se que todas essas estimativas foram significantes de acordo com HPD, observa-se que os valores de L50 para fêmeas variam entre 23,22 cm (modelo *cauchit* - C) e 29,46 cm (modelo potência *probit* - PP). Para os machos, os valores de L50 situam-se entre 22,42 cm (modelo *reversa de potência logit* - RPL) e 27,05 cm (modelo PP). Já os valores de L99, apresentaram maior variabilidade entre os modelos, com estimativas que vão de 31,51 cm (modelo C) a 144,08 cm (modelo PP).

Conforme discutido anteriormente, o modelo com a função de ligação *Potência Cauchit* (PC) apresentou o menor DIC, indicando que este modelo possui um bom ajuste aos dados. Além disso, ele também demonstrou o melhor desempenho na análise do envelope simulado, sugerindo boa adequação aos resíduos. No entanto, sua estimativa para L99 foi alta (133,55 cm para machos e 144,08 cm para fêmeas), o mesmo ocorre na função *reversa de potência Cauchit* (RPC), os valores estimados de L99 foram extremamente elevados, superando 26.000 cm, sugerindo um possível problema de ajuste ou baixa adequação desses dois últimos modelos mencionados (PC e RPC). Esse resultado pode estar relacionado à concentração de amostras em tamanhos menores, conforme observado na distribuição de indivíduos em figura 2.

**Tabela 4:** Valores do comprimento médio para primeira maturação L50, L99, suas respectivas diferenças com intervalos HPD.

Tipo de ligação	Funções de ligação	L50F	L50M	L50F-L50M (HPD)	L99F	L99M	L99F-L99M (HPD)
Básicas	L	25,729	23,654	2,075 (1,98 ; 2,17)	33,579	30,876	2,704 (2,58 ; 2,83)
	P	28,466	26,15	2,316 (2,21 ; 2,42)	28,975	26,617	2,357 (2,25 ; 2,47)
	C	25,226	23,223	1,993 (1,91 ; 2,08)	51,255	47,21	4,045 (3,86 ; 4,23)
Potências	PL	25,707	23,645	2,062 (1,97 ; 2,16)	34,253	31,513	2,741 (1,97 ; 2,16)
	PP	29,465	27,052	2,413 (2,28 ; 2,51)	29,871	27,431	2,44 (2,33 ; 2,56)
	PC	24,785	22,977	1,808 (1,73 ; 1,89)	144,087	133,548	10,538 (9,9 ; 11,14)
Reversas	RPL	24,343	22,426	1,917 (1,83 ; 2,01)	28,229	26,003	2,225 (2,1 ; 2,33)
	RPP	26,375	24,202	2,173 (2,06 ; 2,28)	27,226	24,985	2,241 (2,128 ; 2,24)
	RPC	24,172	22,492	1,68 (1,61 ; 1,75)	26476,883	24631,09	1845,793 (1197,7 ; 2276,41)

As estimativas a posteriori dos parâmetros dos modelos ajustados às funções de ligação indicam variações entre os diferentes métodos utilizados para descrever a maturação do *Prochilodus lineatus* (Apêndice A). Os parâmetros  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  apresentam valores distintos para cada função de ligação, refletindo diferentes padrões de crescimento e maturação.

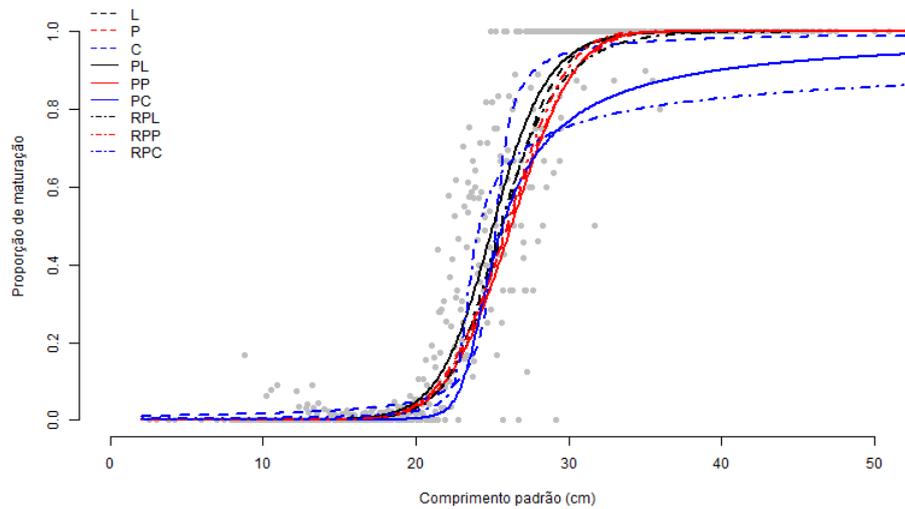
Os desvios padrões das estimativas variam conforme a complexidade da função de ligação, sendo relativamente pequenos para L50 (entre 0,066 e 0,787 cm) e maiores para L99, principalmente em modelos não convencionais como PC e RPC.

As funções *potência logit* (PL) e *reversa de potência logit* (RPL) apresentaram estimativas coerentes com a tradicional *logit*, mas com algumas particularidades. Para a função (PL), o comprimento médio de primeira maturação foi estimado em L50f = 25,705 cm e L50m = 23,648 cm, resultando em uma diferença de 2,057 cm entre os sexos. O valor de L99f foi estimado em 34,253 cm e L99m em 31,513 cm, com uma diferença média de 2,741 cm. Os desvios padrões foram baixos, indicando boa estabilidade das estimativas. O parâmetro de potência  $\lambda = 1,209$ , com um intervalo HPD entre (1,132;1,265), sugere uma leve modificação na curvatura da função em relação ao modelo *logit* e como esperado mais casos imaturos que maduros.

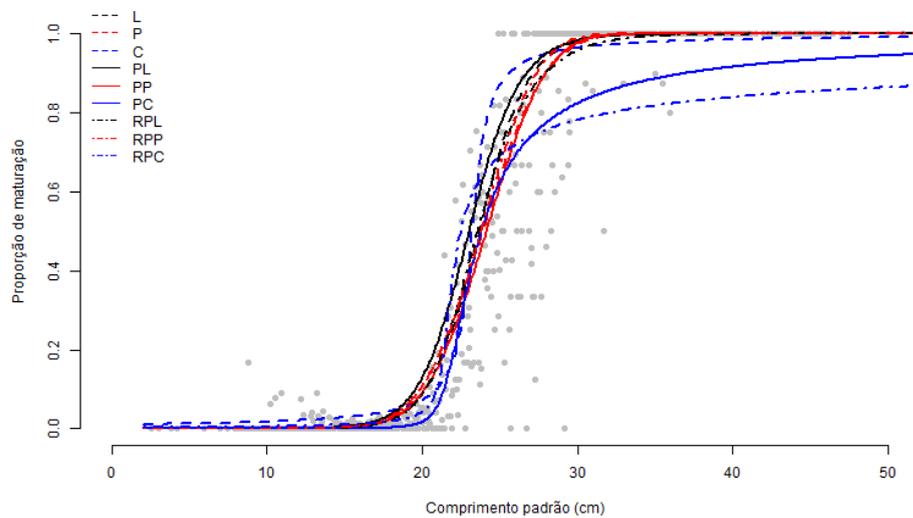
Já para a função (RPL), os valores estimados foram menores, com L50f = 24,351 cm e L50m = 22,431 cm, resultando em uma diferença de 1,919 cm. O valor de L99f foi 28,229 cm e L99m, 26,003 cm, com uma diferença de 2,225 cm. O parâmetro de potência  $\lambda = 0,572$  intervalo HPD (0,525;0,613) indica uma transformação moderada na escala do modelo e também mais casos de imaturos na resposta como esperado do modelo.

No geral, a função PL produziu estimativas próximas às obtidas com o *logit* tradicional, enquanto a RPL resultou em valores um pouco mais baixos para L50 e L99, indicando que essa transformação pode suavizar a inclinação da curva de maturação. Ambas as funções apresentaram estimativas condizentes comparadas a função RPC, cuja estimativa de L99 foi elevada.

Portanto opta-se pelo modelo com função de ligação *reversa de potência logit* (RPL). Esse modelo apresentou um DIC próximo ao do PC (o segundo melhor), um bom ajuste nos envelopes simulados, além de estimar valores de L50 e L99 mais coerentes com a biologia da espécie. O L50 estimado pelo RPL foi 24,34 cm para fêmeas e 22,42 cm para machos, enquanto o L99 foi 28,23 cm para fêmeas e 26,00 cm para machos.



**Figura 3:** Curvas ajustadas da probabilidade de maturação para a fêmea em função do comprimento padrão.



**Figura 4:** Curvas ajustadas da probabilidade de maturação para o macho em função do comprimento padrão.

O gráfico nas figuras 3 e 4 apresentam as curvas de maturação ajustadas por diferentes funções de ligação em relação ao comprimento padrão ( $LS$ ) do peixe *Prochilodus lineatus*, com destaque para os desempenhos das ligações da família *logit*, potência e reversa de potência aplicadas a *cauchit*. As cores indicam os diferentes grupos de modelos: as linhas pretas representam os modelos *logit* (L), *potência logit* (PL) e *reversa de potência logit* (RPL), as vermelhas correspondem aos modelos *probit* (P), *potência probit* (PP) e *reversa de potência probit* (RPP); e as azuis indicam os modelos *cauchit* (C), *potência cauchit* (PC) e *reversa de potência cauchit* (RPC).

Observa-se que os modelos baseados na ligação *logit* (linhas pretas) apresentam um ajuste consistente, capturando a transição entre indivíduos imaturos e maduros. As

curvas *logit* e suas variações de potência e reversa de potência mostram inclinações semelhantes, sugerindo que essas funções são adequadas para descrever o processo de maturação. As linhas vermelhas, correspondentes às versões *probit*, também apresentam um ajuste semelhante ao *logit*, com pequenas variações na inclinação e nas caudas da distribuição.

Por outro lado, as funções baseadas na ligação *cauchit* (linhas azuis) apresentam um ajuste visivelmente inferior, especialmente as versões *potência cauchit* e *reversa de potência cauchit*. Essas curvas mostram maior dispersão nas extremidades, com um crescimento mais lento na transição para a maturação e uma cauda superior prolongada, indicando que esses modelos não capturam o comportamento final dos dados. Em especial, a função *reversa de potência cauchit* comparada com a curva ajustada dos dados observados (pontos cinza), nota-se que a *reversa de potência cauchit* não acompanha bem a distribuição.

Essas diferenças sugerem que, entre os modelos avaliados, as funções baseadas na família *logit* são mais adequadas para estimar parâmetros como L50 e L99. Já os modelos baseados na ligação *cauchit* apresentam um desempenho inferior, especialmente quando combinados com funções de potência e reversa de potência, evidenciando um ajuste inadequado aos dados observados. Dessa forma, a escolha do modelo deve levar em consideração não apenas o critério de informação (*DIC*), mas também a coerência biológica e estatística dos ajustes.

#### 4. DISCUSSÃO

Os resultados desta pesquisa indicam que a população analisada de *Prochilodus lineatus* apresentou um desbalanceamento na maturação, com 88,62% dos indivíduos classificados como imaturos e apenas 11,38% como maduros, além de uma alta concentração de amostras em comprimentos inferiores a 20 cm. Todos os modelos ajustados passaram nos testes de convergência, e a avaliação do *Deviance Information Criterion (DIC)* destacou o modelo *potência cauchit (PC)* como o mais parcimonioso, seguido pelo *reversa de potência logit (RPL)*. No entanto, apesar do menor *DIC*, o modelo *PC* apresentou estimativas de L99 biologicamente incoerentes, assim como sua versão reversa (*RPC*), comprometendo sua aplicabilidade.

A análise das curvas ajustadas mostrou que os modelos baseados na família *logit* apresentaram melhor ajuste aos dados observados, enquanto os modelos da família *cauchit* exibiram desempenho inferior, com caudas prolongadas e menor aderência aos dados observados. Considerando a qualidade do ajuste, parcimônia e coerência biológica das estimativas, o modelo *RPL* foi o mais adequado para descrever a maturação da espécie, fornecendo estimativas de L50 (24,34 cm para fêmeas e 22,42 cm para machos) e L99 (28,23 cm para fêmeas e 26,00 cm para machos), além de manter um equilíbrio entre sensibilidade e especificidade na classificação dos indivíduos.

Embora a regressão logística seja amplamente utilizada por sua facilidade de interpretação, o foco deste estudo está nos parâmetros L50 e L99, que possuem aplicações diretas na biologia pesqueira e na formulação de políticas de manejo. Os resultados demonstram que a escolha da função de ligação influencia essas estimativas, sendo essencial considerar não apenas critérios estatísticos, mas também a coerência biológica dos valores obtidos. Algumas funções conseguem capturar melhor a distribuição dos dados, permitindo um ajuste mais adequado à realidade biológica da espécie.

Em estudos anteriores, como o de Alves et al. (2020), a ligação *logit* foi utilizada devido à sua praticidade na interpretação dos parâmetros. Ao analisar as métricas de qualidade de ajuste e desempenho dos modelos (*DIC*, acurácia, sensibilidade, especificidade, curvas ajustadas e tempo de processamento), verificou-se que o modelo *PC* teve o melhor ajuste aos dados, enquanto o *cauchy* foi o menos adequado. Percebe-se que a incorporação da potência na ligação possui maior flexibilidade ao modelo. No entanto, o modelo *PC* apresentou estimativas incoerentes de *L99*, comprometendo sua aplicabilidade. Modelos com ligações do tipo *probit* mostraram-se mais sensíveis, enquanto os modelos com ligações do tipo *cauchy* apresentaram maior sensibilidade a variações nos dados. Os modelos que envolveram a ligação *logit* demonstraram um equilíbrio entre sensibilidade e especificidade, além de manter níveis equivalentes de acurácia. Como o modelo *RPL* foi o segundo com o menor *DIC*, optou-se por escolhê-lo como o melhor modelo entre os analisados.

Outros estudos que modelam respostas binárias, como os de Bazán et al. (2017), Anyosa (2017) e Galo (2020), também utilizaram funções de ligação simétricas e assimétricas, evidenciando que estas últimas tendem a proporcionar um melhor ajuste. Esse fato pode estar relacionado à capacidade das ligações assimétricas de capturar o desbalanceamento da resposta, ainda que sua assimetria seja fixa, o que impede uma flexibilidade total na representação da assimetria real da resposta (Abanto-Valle, Bazán e Smith, 2014).

A metodologia *Bayesiana*, utilizada para obter as estimativas dos parâmetros, mostrou-se equivalente à abordagem *frequentista* neste estudo, uma vez que foram adotadas distribuições a priori não informativas e uma amostra grande o suficiente para garantir boas inferências sem recorrer a teorias assintóticas (Paal B. van der, 2014; Fernandes, 2016). A abordagem *Bayesiana* é vantajosa por permitir maior flexibilidade na escolha dos estimadores, como demonstrado neste estudo, no qual a média e a mediana se mostraram equivalentes para a obtenção da máxima informação. Ademais, quando existe conhecimento prévio sobre os parâmetros em estudo, ele pode ser incorporado no processo de inferência (Paulino, 2003).

Na área das ciências da vida, muitas pesquisas apresentam distribuições desbalanceadas, como no caso da maturação de peixes (Barbosa et al., 2022). Considerando as implicações da maturação reprodutiva precoce ou tardia para o recrutamento e a sustentabilidade das populações pesqueiras (Silva et al., 2021),

compreender os fatores que influenciam esse processo é essencial. Estudos indicam que variáveis como sexo e comprimento padrão desempenham um papel crucial na modelagem da probabilidade de maturação (Barbosa et al., 2022), permitindo a obtenção de estimativas importantes, como *L50* (comprimento em que 50% da população está madura) e *L99* (comprimento em que 99% está madura) (Silva et al., 2021). Essas estimativas são fundamentais para identificar padrões de maturação e avaliar o desempenho de diferentes funções de ligação. O uso de diferentes funções de ligação nessas investigações fornece uma base para decisões em gestão e conservação das espécies.

## 5. CONSIDERAÇÕES FINAIS

Neste trabalho, foram revisadas, além das funções de ligação tradicionais para regressão binária, as ligações potência e reversa de potência, definidas por Abanto-Valle, Bazán e Smith (2014), Bazán, Romeo e Rodrigues (2014) e Bazán et al. (2017), as quais foram obtidas a partir da generalização da distribuição normal proposta por Gupta e Gupta (2008) e pela adequação na propriedade de reversibilidade. Aqui, as distribuições de base foram baseadas nas funções de distribuição acumulada simétricas (Logística, Normal e Cauchy) e assimétricas (Gumbel de valor Mínimo e Máximo), assim como Bazán et al. (2017) e Anyosa (2017) o fizeram.

Tais funções de ligação apresentam-se como alternativa para respostas binárias quando estas são desbalanceadas, já que possuem um parâmetro relacionado à assimetria ( $\lambda$ ). Além disso, a estimação dos parâmetros a partir da abordagem bayesiana utilizando o processo iterativo MCMC mostrou-se uma opção prática para a implementação dos modelos com as ligações propostas, além de ser uma abordagem flexível na escolha do estimador para os parâmetros. Os códigos para todos os modelos foram desenvolvidos no programa R e alguns exemplos são apresentados no Apêndice C.

A partir da aplicação dessas metodologias à modelagem da maturação do *Prochilodus lineatus*, foi possível avaliar o desempenho das diferentes funções de ligação na estimação dos comprimentos de maturação *L50* e *L99*, bem como comparar os modelos por meio do DIC, análise de resíduos (envelope simulado), avaliação preditiva, análise gráfica. Os resultados indicaram que a função de ligação *potência cauchit* (PC) apresentou o melhor DIC e o melhor ajuste no envelope simulado, contudo, superestimou o *L99*, possivelmente devido à alta concentração de amostras nos menores comprimentos. Diante desse comportamento, o modelo com função de ligação *reversa potência logit* (RPL) foi considerado o mais adequado, pois proporcionou um bom equilíbrio entre as estimativas de *L50* e *L99*, também equilíbrio entre a avaliação preditiva, além de apresentar o segundo melhor DIC, um envelope simulado e tempo computacional razoável.

Neste estudo, mostrou-se que a escolha da função pode influenciar as estimativas biológicas. A abordagem utilizada contribui para a metodologias estatísticas aplicadas à biologia pesqueira, oferecendo um referencial para futuras análises de maturação em *Prochilodus lineatus* e outras espécies. Além disso, os resultados podem ser úteis para a formulação de políticas de manejo sustentável, uma vez que a correta determinação do *L50* e *L99* é fundamental para a definição de tamanhos mínimos de captura e estratégias

de conservação das populações naturais.

Entretanto, algumas limitações podem ser observadas ao longo do estudo. A concentração desbalanceada de indivíduos imaturos pode ter influenciado a estimação dos parâmetros, em especial do L99, impactando a performance dos modelos. Além disso, a análise foi baseada em dados binários de maturação, sem considerar possíveis efeitos ambientais ou variações sazonais que podem influenciar o processo de maturação dos indivíduos. Estudos futuros podem explorar abordagens que incorporem essas variáveis adicionais, bem como testar novas funções de ligação menos convencionais.

Por fim, recomenda-se a continuidade da investigação sobre a modelagem da maturação, especialmente via abordagens bayesianas com efeitos aleatórios, que podem capturar variações individuais entre populações ou entre diferentes períodos de coleta. Além disso, aprimorar o desenho amostral, garantindo uma distribuição mais equilibrada entre indivíduos maduros e imaturos, pode minimizar possíveis vieses nas estimativas dos parâmetros.

De forma geral, as funções de ligação potência e reversa de potência apresentam-se como alternativas para a modelagem de dados binários desbalanceados em distintos campos de estudo, incluindo a biologia pesqueira, como neste trabalho. Como proposta futura, tais modelos podem ser disponibilizados em um pacote do R, utilizando a integração com o software JAGS, facilitando sua aplicação em diferentes contextos científicos e ampliando sua acessibilidade para pesquisadores da área.

## REFERÊNCIAS

AGOSTINHO, AA, Gomes, LC, Veríssimo, S. *et al.* Regime de cheias, regulação de barragens e peixes no Alto Rio Paraná: efeitos sobre atributos de assembleia, reprodução e recrutamento. *Rev Fish Biol Fisheries* **14**, 11–19 (2004). <https://doi.org/10.1007/s11160-004-3551-y>

ABANTO-VALLE, C. A.; BAZÁN, J. L.; SMITH, A. C. *State space mixed models for binary responses with skewed inverse links using JAGS*. Rio de Janeiro, 2014.

AGRESTI, A. **An Introduction to Categorical Data Analysis**: John Wiley & Sons, (2007).

ALBERT, J. H.; CHIB, S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, Informa UK Limited, v. 88, n. 422, p. 669–679, 1993. doi: <<https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1993.10476321>>.

ANYOSA, S. A. C. *Regressão binária usando ligações potência e reversa de potência*. Dissertação (Mestrado em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística) — Instituto de Ciências Matemáticas e de Computação, Universidade Federal de São Paulo, São Carlos, 2017.

ARANDA-ORDAZ, F. J. On two families of transformations to additivity for binary response data. *Biometrika*, Oxford University Press (OUP), v. 68, n. 2, p. 357–363, 1981. doi: <<https://doi.org/10.1093/biomet/68.2.357>>.

ATKINSON, A. C. Two graphical displays for outlying and influential observations in regression. *Biometrika*, Oxford University Press, v. 68, n. 1, p. 13–20, 1981. doi: <<https://doi.org/10.1093/biomet/68.1.13>>.

BAZÁN, J. *et al.* Power and reversal power links for binary regressions: An application for motor insurance policyholders. *Applied Stochastic Models in Business and Industry*, Wiley, v. 33, n. 1, p. 22–34, 2017. doi: <<https://doi.org/10.1002/asmb.2215>>.

BARBOSA, L. M.; COSTA, R. M. R.; MUNIZ, C. C.; MATEUS, L. A. F. Tamanho da primeira maturação de duas espécies de interesse pesqueiro na bacia do Alto Rio

Paraguai – Mato Grosso: subsídios para o estabelecimento do tamanho mínimo de captura. *Revista Biodiversidade*, v. 21, n. 2, p. 46-57, 2022. Disponível em: [periodicoscientificos.ufmt.br](http://periodicoscientificos.ufmt.br).

BAZÁN, J. L.; ROMEO, J. S.; RODRIGUES, J. Bayesian skew-probit regression for binary response data. *Brazilian Journal of Probability and Statistics*, Institute of Mathematical Statistics, v. 28, n. 4, p. 467–482, 2014. doi: <<https://doi.org/10.1214/13-bjps218>>.

BAZÁN, J.; TORRES-AVILÉS, F.; SUZUKI, A.; LOUZADA, F. Power and reversal power links for binary regressions: Na application for motor insurance policy holders. **Applied Stochastic Models in Business and Industry**, John Wiley & Sons, Ltd, p.n/a-n/a, 2016.

BRESLOW, N.E.; CLAYTON, D.G. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.*, v.88, p.9-25, 1993.

BROWN-PETERSON, Nancy J. *et al.* A standardized terminology for describing reproductive development in fishes. **Marine and Coastal Fisheries**, v. 3, n. 1, p. 52-70, 2011.

CAMERON, A. C., & TRIVEDI, P. K. **Regression Analysis of Count Data**. Cambridge University Press, (2013).

CORRÊA ALVES, Diego *et al.* New approaches to basic population ecology studies: Revealing more complex patterns of a small Characidae that inhabit streams. **Ecology of Freshwater Fish**, v. 29, n. 4, p. 574-587, 2020.

CHEN, M.-H.; DEY, D. K.; SHAO, Q.-M. A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, Informa UK Limited, v. 94, n. 448, p. 1172–1186, 1999. doi: <<https://doi.org/10.1080/01621459.1999.10473872>>.

CZADO, C.; SANTNER, T. J. The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference*, Elsevier BV, v. 33, n. 2, p. 213–231, 1992. doi: <[https://doi.org/10.1016/0378-3758\(92\)90069-5](https://doi.org/10.1016/0378-3758(92)90069-5)>.

DA SILVA, Andson Nunes; ANYOSA, Susan; BAZAN, Jorge Luis. Modelagem Bayesiano de regressão binária para dados desbalanceados usando novas ligações. **Brazilian**

**Journal of Biometrics**, v. 38, n. 4, p. 385-417, 2020.

DEMETRIO, C.G.B. Modelos lineares generalizados em experimentação agrônômica. Piracicaba: ESALQ, 2001. 113p.

DOBSON, A. *An introduction to generalized linear models*. Boca Raton: Chapman & Hall/CRC, 2002. ISBN 1-58488-165-8.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, v. 5, n. 3, p. 236–244, 1996. doi: <<https://doi.org/10.1080/10618600.1996.10474708>>.

FERNANDES, AMC. Regressão binária nas abordagens clássica e bayesiana [dissertação de mestrado]. São Carlos: Universidade Federal de São Paulo (USP); 2016.

GARCÍA, D.; ROMERO, N.; CRESPO, S. Modelling fish reproductive patterns by using generalized linear models: logistic regression and probit models. *Fisheries Research*, v. 82, n. 3, p. 255-261, 2006.

GALO, Rafaela et al. Bayesian binary regression using power and power reverse link functions: an application to premature birth data. **Brazilian Journal of Biometrics**, v. 41, n. 2, p. 131-143, 2023.

GALO, R. *Modelagem bayesiana para dados de nascimentos prematuros desbalanceados*. Dissertação (Mestrado em Bioestatística - Programa Pós-Graduação em Bioestatística) — Universidade Estadual de Maringá, Maringá, 2020.

GEWEKE, John. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. **Bayesian statistics**, v. 4, p. 169-193, 1992.

GELMAN, Andrew *et al.* **Bayesian data analysis**. Chapman and Hall/CRC, 2013.

GIOLO, S. R. *Introdução à análise de dados categóricos com aplicações*. [S.l.]: Editora Blucher, 2017. ISBN 9788521211877.

GRAÇA, W. & CS PAVANELLI, 2007. Peixes da planta de inundação do alto rio Paraná e áreas adjacentes. EDUEM, Maringá.

GUPTA, R. D.; GUPTA, R. C. Analyzing skewed data by power normal model. *Test*, Springer Science and Business Media LLC, v. 17, n. 1, p. 197–210, 2008. doi: <<https://doi.org/10.1007/s11749-006-0030-x>>.

HEIDELBERGER, Philip; WELCH, Peter D. **Simulation Run Length Control in the Presence of an Initial Transient**. *Operations Research*, v. 31, n. 6, p. 1109-1144, 1983. Disponível em: <<https://doi.org/10.1287/opre.31.6.1109>>. Acesso em: 9 de ago. 2024.

HUAYANAY, A. d. L. C. *Modelos de regressão para resposta binária na presença de dados desbalanceados*. Dissertação (Mestrado em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística) — Instituto de Ciências Matemáticas e de Computação, Universidade Federal de São Paulo, São Carlos, 2019.

JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. [S.1.]: John Wiley & Sons, 2013. V. 398.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNANN, N. *Continuous univariate distributions*. 2nd edn. New York: Wiley, 1994. ISBN 978-0-471-58495-7.

KING, Michael. **Fisheries biology, assessment and management**. John Wiley & Sons, 2007.

LEMONTE, A. J.; BAZÁN, J. L. New links for binary regression: an application to coca cultivation in peru. *Test*, Springer, v. 27, n. 3, p. 597–617, 2018. doi: <<https://doi.org/10.1007/s11749-017-0563-1>>.

MCCULLAGH, P.; NELDER, J. *Generalized linear models*. 2. ed. Boca Raton London New York: Chapman and Hall, 1989. ISBN 0-412-31760-5.

MORGAN, M. Joanne. Integrating reproductive biology into scientific advice for fisheries management. **Journal of Northwest Atlantic Fishery Science**, v. 41, 2008.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, JSTOR, v. 135, n. 3, p. 370, 1972. doi:<<https://doi.org/10.2307/2344614>>.

NUNES, J. A. R.; MORAIS, A. R.; BUENO FILHO, J. S. S. Modelagem da superdispersão em dados por um modelo linear generalizado misto. *Rev. Mat. Estat.*, v.22, p.55-70, 2004.

PAULINO, C. D. M.; TURKMAN, M. A. A.; MURTEIRA, B. *Estatística bayesiana*. Lisboa:Fundação Calouste Gulbenkia, 2003. ISBN 972-31-1043-1.

PAULA, G. A. *odelos de regressão: com apoio computacional*. 2013. Disponível em: <[https://www.ime.usp.br/~giapaula/texto\\_2013.pdf](https://www.ime.usp.br/~giapaula/texto_2013.pdf)>. Acesso em: 12/10/2024.

PLUMMER, Martyn. rjags: Bayesian Graphical Models using MCMC. R package version 4-12, 2023. Disponível em: <<https://cran.r-project.org/web/packages/rjags/index.html>>. Acesso em: 19 ago. 2024.

PLUMMER, Martyn *et al.* **Coda: Convergence Diagnosis and Output Analysis for MCMC**. *R News*, v. 6, n. 1, p. 7-11, 2006. Disponível em: <[https://CRAN.R-project.org/doc/Rnews/Rnews\\_2006-1.pdf](https://CRAN.R-project.org/doc/Rnews/Rnews_2006-1.pdf)>. Acesso em: 19 de ago. 2024.

R CORE TEAM, R: A Language and Environment for Statistical Computing. Vienna, Austria. 2024.

SILVA, Á. P. C.; CARVALHO, I. F. S.; DINIZ, A. L. C.; CANTANHÊDE, L. G.; ALMEIDA, Z. S. Tamanho de primeira maturação para cinco espécies de peixes em uma área da Costa Amazônica Maranhense, Brasil. *Research, Society and Development*, v. 10, n. 3, e23610313146, 2021. Disponível em: [rsdjournal.org](https://rsdjournal.org).

SOUZA FILHO, EE & JC STEVAUX, 2004. Geologia e geomorfologia do complexo fluvial Baía–Curutuba–Ivinhema. Em Thomaz, SM, AA Agostinho & NS Hahn (eds), *O Alto Rio Paraná e Sua Várzea: Aspectos Físicos, Ecologia e Conservação*. Editores Backhuys, Leiden: 1–29.

SPIEGELHALTER, David J. *et al.* Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 64, n. 4, p. 583-639, 2002. Disponível em: <<https://doi.org/10.1111/1467-9868.00353>>. Acesso em: 19 de ago. 2024.

SU, Yu-Sung; YAJIMA, Masanao. **R2jags: Using R to Run 'JAGS'**. 2015. Disponível em: <<https://CRAN.R-project.org/package=R2jags>>. Acesso em: 9 de ago. 2024.

VAN DER PAAL, Bart. A comparison of different methods for modelling rare events data. **Master of Statistical Data Analysis thesis, Universiteit Gent**, 2014.

VENABLES WN. Modern applied statistics with S. 4th ed. New York: Springer, 2002.

VASCONCELOS, L. P.; ALVES, D. C.; GOMES, L. C. Spatial and temporal variations among fish with similar strategies: patterns of reproductive guilds in a floodplain. **Hydrobiologia**, v. 726, p. 213-228, 2014.

VAZZOLER, AEA de M. Biologia da reproducao de peixes teleosteos: teoria e pratica: Maringa. **Sao Pablo, Brazil: EDUEM**, 1996.

## APÊNDICES

### APÊNDICE A

#### Fundamentação teórica

Os Modelos Lineares Generalizados (MLG) introduzidos por Nelder e Wedderburn (1972) é uma técnica estatística amplamente utilizada para examinar a relação entre variáveis independentes, chamada de explicativas, e uma variável de resposta, cuja distribuição de probabilidade pertence à família exponencial de distribuições, as quais possuem propriedades assintóticas favoráveis para inferências precisas, ou seja, para grandes amostras, as estimativas e os testes estatísticos baseados nessa distribuição são confiáveis e se aproximam dos valores verdadeiros, também proporcionam maior flexibilidade à modelagem (DOBSON, 2002).

Quando a resposta relacionada ao sucesso ou fracasso de um evento é binária, o MLG empregado é o modelo de regressão Binomial. Dessa forma, busca-se modelar a probabilidade da variável resposta ser um sucesso considerando as variáveis independentes por meio de uma função de ligação apropriada, neste caso geralmente é utilizada a ligação *logit*.

#### Regressão binária

A regressão binária é uma técnica estatística usada para modelar a relação entre uma variável dependente binária,  $Y$ , e um conjunto de variáveis independentes  $(X_1, X_2, \dots, X_p)$ . Dado um conjunto de  $n$  observações, com  $Y_i$  sendo a variável binária dependente e  $X = (1, X_1, X_2, \dots, X_p)$  às variáveis independentes para a  $i$ -ésima observação, o vetor de coeficientes  $\beta = (\beta_0, \beta_1, \dots, \beta_p)$  descreve a influência dessas variáveis explicativas. Considere  $Y_i = 1$  com probabilidade  $\theta_i$  e  $Y_i = 0$  com probabilidade  $1 - \theta$ , assim o modelo de regressão binária é dado por:

$$Y_i \sim \text{Bernoulli}(\theta_i)$$

$$\theta_i = F(\eta_i) = F(X\beta),$$

Sendo que,  $F(\cdot)$  denota a função de distribuição acumulada (fda),  $F^{-1}(\cdot)$  denomina-se função de ligação e  $\eta_i = X\beta$  é o  $i$ -ésimo preditor linear.

De acordo com ANYOSA (2017), a função de ligação  $F^{-1}(\cdot)$  transforma o suporte de  $p$  no conjunto dos números reais. Quando o preditor linear  $\eta$  com valores reais são avaliados na fda,  $F(\cdot)$ , os valores obtidos são garantidos como valores de probabilidades válidos que ficam restritos entre 0 e 1.

Em modelos lineares generalizados, usa-se uma função de ligação quando

estiver realizando uma análise estatística que envolva a relação entre variáveis dependentes e independentes em um modelo de regressão. A importância das funções de ligação reside no fato de fornecer uma maneira de modelar e entender como as variáveis independentes influenciam a variável dependente em um contexto de regressão. Isto auxilia para previsões e também para relação entre variáveis em um conjunto de dados.

Abaixo estão algumas funções de ligação clássicas para dados binários.

- Função Logit: Utilizada em modelos de regressão logística, onde o logaritmo da razão de chances é modelado como uma combinação linear das variáveis independentes (CAMERON & TRIVEDI, 2013).
- Função Probit: Similar à função logit, mas usa a distribuição normal acumulada inversa para modelar a probabilidade de um evento binário (Agresti, 2007).
- Função Cauchit: A função Cauchit é utilizada em modelos de regressão logística para lidar com valores extremos e distribuições de probabilidade que não são bem aproximadas pela distribuição normal. Ela suaviza as estimativas e reduz o impacto de valores extremos nos resultados do modelo. (HOSMER JR. *et al.*, 2013).
- Função Log-log: Usada para modelar a taxa de falha em análises de sobrevivência, onde os dados são censurados e não censurados (CAMERON & TRIVEDI, 2013).
- Função Complemento Log-log: Também usada em análises de sobrevivência, especialmente quando há uma taxa de falha muito baixa (CAMERON & TRIVEDI, 2013).

Na Tabela A1 tem-se as funções de ligações comumente utilizadas e suas respectivas inversas:

**Tabela A1:** Funções de ligações clássicas.

Origem	Nome	F(.)	F <sup>-1</sup> (.)
Logística	<b>Logit (L)</b>	$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$	$\log\left(\frac{p_i}{1 - p_i}\right) = \eta_i$
Normal	<b>Probit (P)</b>	$p_i = \Phi(\eta_i)$	$\Phi^{-1}(p_i) = \eta_i$
Cauchy	<b>Cauchit (C)</b>	$p_i = 0,5 + \frac{\arctan(\eta_i)}{\pi}$	$\tan(\pi(p_i - 0,5)) = \eta_i$
Gumbel de Valor mínimo	<b>Cloglog (CLL)</b>	$p_i = 1 - \exp(-\exp(\eta_i))$	$\log(-\log(1 - p_i)) = \eta_i$
Gumbel de valor máximo	<b>Loglog (LL)</b>	$p_i = \exp(-\exp(-\eta_i))$	$-\log(-\log(p_i)) = \eta_i$

Fonte: Adaptado de Anyosa (2017).

A ligação *logit* é obtida da f.d.a da distribuição Logística, a *probit* da Normal, a *cauchit* da Cauchy, a *cloglog* da Gumbel de Valor Mínimo e a *loglog* da Gumbel de Valor Máximo.

## Propriedades de simetria e reversibilidade de distribuições de probabilidade

Bazán *et al.* (2016) definem as propriedades necessárias a serem consideradas quando são usadas diferentes funções de ligação. As propriedades de simetria e reversibilidade, relacionadas a uma distribuição de probabilidade qualquer, são definidas como:

**Definição (Propriedade de simetria):** Seja  $T$  uma variável aleatória (v.a.) com uma determinada distribuição de probabilidade contínua. Considere também que  $F(\cdot)$  denote a função distribuição acumulada (f.d.a.) e  $f(\cdot)$  a função densidade de probabilidade (f.d.p) desta distribuição. A v.a.  $T$  pode ser denotada por  $T \sim F(\cdot)$  ou  $f(\cdot)$ , isto é,  $T$  tem distribuição de probabilidade com f.d.a igual a  $F(\cdot)$  ou f.d.p igual a  $f(\cdot)$ .

Uma distribuição de probabilidade é considerada como simétrica se, para um valor  $x_0$  for verificado que  $f(x_0 + \epsilon) = f(x_0 - \epsilon)$  para todo número real  $\epsilon$ . Caso a distribuição não satisfaça essa propriedade, é dita assimétrica. Se uma distribuição é simétrica em torno de  $x_0$ , este é dito ponto de simetria. Além disso, a mediana e a média (se existirem) ocorrem neste ponto.

**Definição (Propriedade de reversibilidade)** A distribuição de probabilidade da v.a.  $T$  apresenta a propriedade de reversibilidade se a v.a.  $-T$  apresentar f.d.a.  $G(\cdot) = 1 - F(-\cdot)$ , em que  $F(\cdot)$  é a f.d.a. de  $T$ , isto é,  $-T \sim G(\cdot) \equiv 1 - F(-\cdot)$ , tal que  $F(-\cdot)$  é usada para calcular a f.d.a. de  $-T$ , refletindo a distribuição de  $T$  em torno de zero.

Para verificar se tal propriedade é válida em uma distribuição de probabilidade, a Equação abaixo deve ser verdadeira.

$$F(\cdot) + G(-\cdot) = 1, \quad (2.1)$$

em que  $F(\cdot)$  é a f.d.a. da distribuição de probabilidade de interesse e  $G(-\cdot)$  a f.d.a. da respectiva distribuição de probabilidade reversa.

Graficamente, as f.d.a. da distribuição probabilidade e sua reversa, apresentam-se uma como um espelho da outra (ANYOSA, 2017).

## Função de ligação de potência e potência reversa

A partir da generalização da distribuição normal sugerida por Gupta e Gupta (2008), Abanto-Valle, Bazán e Smith (2014) mostram que a distribuição de probabilidade potência é adquirida elevando-se a uma potência positiva ( $\lambda$ ) a f.d.a. de qualquer distribuição de probabilidade contínua nos reais, unimodal e log côncava. A partir disso,

uma nova distribuição de probabilidade assimétrica é determinada com um parâmetro ( $\lambda$ ) associado à assimetria que os dados analisados apresentam.

Como as distribuições de probabilidade potência satisfazem a propriedade de reversibilidade, é possível obter as respectivas distribuições de probabilidade reversa potência que também apresentam o parâmetro de assimetria ( $\lambda$ ) (BAZÁN *et al.*, 2016). A seguir as definições para estas distribuições:

**Definição (Distribuição Potência – P)** Uma v.a. univariada  $T$  tem distribuição de probabilidade potência, isto é,  $T \sim P(\mu, \sigma^2, \lambda)$ , sendo  $\mu \in \mathbb{R}$  parâmetro de locação,  $\sigma^2 > 0$  parâmetro de escala e  $\lambda > 0$  parâmetro de forma, sua f.d.p. e f.d.a. são, respectivamente, da forma:

$$f_P(t|\mu, \sigma^2, \lambda) = \frac{\lambda}{\sigma} g\left(\frac{t-\mu}{\sigma}\right) \left[G\left(\frac{t-\mu}{\sigma}\right)\right]^{\lambda-1}, \quad t \in \mathbb{R} \quad (2.2)$$

e

$$F_P(t|\mu, \sigma^2, \lambda) = \left[G\left(\frac{t-\mu}{\sigma}\right)\right]^\lambda, \quad t \in \mathbb{R}, \quad (2.3)$$

sendo  $g(\cdot)$  uma f.d.p. de característica unimodal, log côncava com suporte nos reais e  $G(\cdot)$ , a correspondente f.d.a. de qualquer distribuição de probabilidade podendo ser simétrica ou assimétrica, que será chamada de distribuição base. Quando  $\lambda = 1$  nas equações acima,  $T$  tem a mesma distribuição de probabilidade que a distribuição base.

**Definição (Distribuição Reversa de Potência - RP)** Uma v.a. univariada  $T$  com distribuição de probabilidade reversa de potência, isto é,  $T \sim RP(\mu, \sigma^2, \lambda)$ , sendo  $\mu \in \mathbb{R}$  parâmetro de locação,  $\sigma^2 > 0$  parâmetro de escala e  $\lambda > 0$  parâmetro de forma, se sua f.d.p. e f.d.a. são, respectivamente, da forma:

$$f_{RP}(t|\mu, \sigma^2, \lambda) = \frac{\lambda}{\sigma} g\left(-\frac{t-\mu}{\sigma}\right) \left[G\left(-\frac{t-\mu}{\sigma}\right)\right]^{\lambda-1}, \quad t \in \mathbb{R} \quad (2.4)$$

e

$$F_{RP}(t|\mu, \sigma^2, \lambda) = 1 - \left[G\left(-\frac{t-\mu}{\sigma}\right)\right]^\lambda, \quad t \in \mathbb{R}, \quad (2.5)$$

sendo  $g(\cdot)$  uma f.d.p. de característica unimodal, log côncava com suporte nos reais e  $G(\cdot)$ , a correspondente f.d.a. de qualquer distribuição de probabilidade podendo ser simétrica ou assimétrica, que será chamada de distribuição base. Quando  $\lambda = 1$  nas equações (2.4) e (2.5),  $T$  tem a mesma distribuição de probabilidade que a distribuição base.

As distribuições potência e reversa de potência apresentam sua forma padrão quando o parâmetro de locação  $\mu = 0$  e o de escala  $\sigma^2 = 1$ . Assim, uma v.a.  $W$  tem distribuição potência padrão, isto é,  $W \sim P(\lambda)$ , se a f.d.a. e a f.d.p são, respectivamente, dadas por:

$$f_P(w|\lambda) = \lambda g(w)[G(w)]^{\lambda-1} \text{ e } F_P(w|\lambda) = [G(w)]^\lambda. \quad (2.6)$$

Além disso, uma v.a.  $W$  tem distribuição de probabilidade reversa de potência padrão, isto é,  $W \sim RP(\lambda)$ , se a f.d.p. e a f.d.a são, respectivamente, dadas por:

$$f_{RP}(w|\lambda) = \lambda g(w)[G(-w)]^{\lambda-1} \text{ e } F_{RP}(w|\lambda) = 1 - [G(-w)]^\lambda. \quad (2.7)$$

Para evitar confusões, as distribuições potências e reversas de potência em sua forma padrão passam a ser denominadas apenas como distribuição potência e distribuição reversa de potência. Portanto, é possível obter uma variedade de ligações assimétricas para a regressão binária que apresentam o parâmetro  $\lambda$  associado à assimetria dos dados.

### Regressão binária utilizando funções de potência e potência reversa

Anyosa (2017) propõem utilizar as distribuições potência e reversa de potência formadas das distribuições básicas apresentadas na (Tabela A1) para obter novas funções de ligação binária com característica assimétrica. Assim, os modelos de regressão binária obtidos destas distribuições são dados na forma descrita na (Tabela 2), sendo  $F(\cdot) = F_\lambda(\cdot)$  a f.d.a. de alguma distribuição potência ou reversa de potência,  $F^{-1} = F_\lambda^{-1}(\cdot)$  a função de ligação potência ou reversa de potência.

A Tabela A2 apresenta as distribuições de probabilidade que definem as funções de ligação assimétricas para a análise de dados binários, a f.d.a  $F_\lambda(\cdot)$  e a função de ligação  $F_\lambda^{-1}(\cdot)$ .

Na literatura, pode-se verificar que a partir das Definições de potência e potência reversa foram obtidas: a distribuição Potência Normal (PN), proposta por Gupta e Gupta (2008); a distribuição Reversa de Potência Normal (RPN) por Bazán, Romeo e Rodrigues (2014); as distribuições Potência Gumbel de Valor Máximo (PGVMa) e Potência Gumbel de Valor Mínimo (PGVMi), propostas por Abanto-Valle, Bazán e Smith (2014); e as distribuições Potência Cauchy (PC), Reversa de Potência Cauchy (RPC), Reversa de Potência Gumbel de Valor Máximo (RPGVMa) e Reversa de Potência Gumbel de Valor Mínimo (RPGVMi) por Bazán *et al.* (2017).

No entanto, as distribuições Reversa de Potência Logística (RPL) e Potência Logística (PL) não se apresentaram das definições (2.3) e (2.4). A distribuição RPL foi proposta por Nagler (1994) e foi denominada de *Scobit*, já a distribuição PL foi obtida por uma generalização da distribuição logística chamada de generalização do tipo Johnson, Kotz e Balakrishnan (1994).

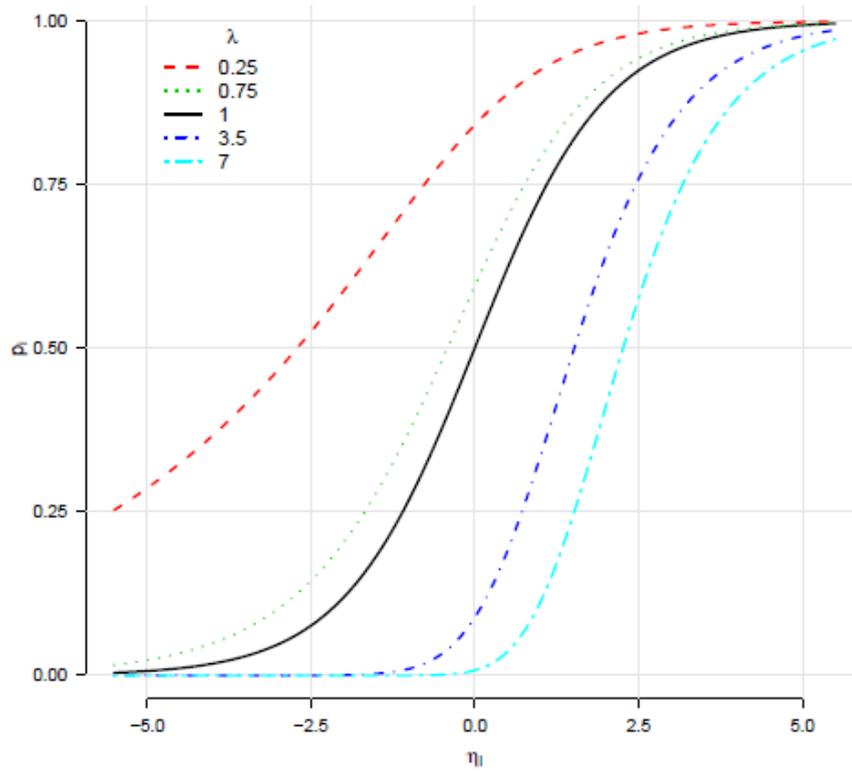
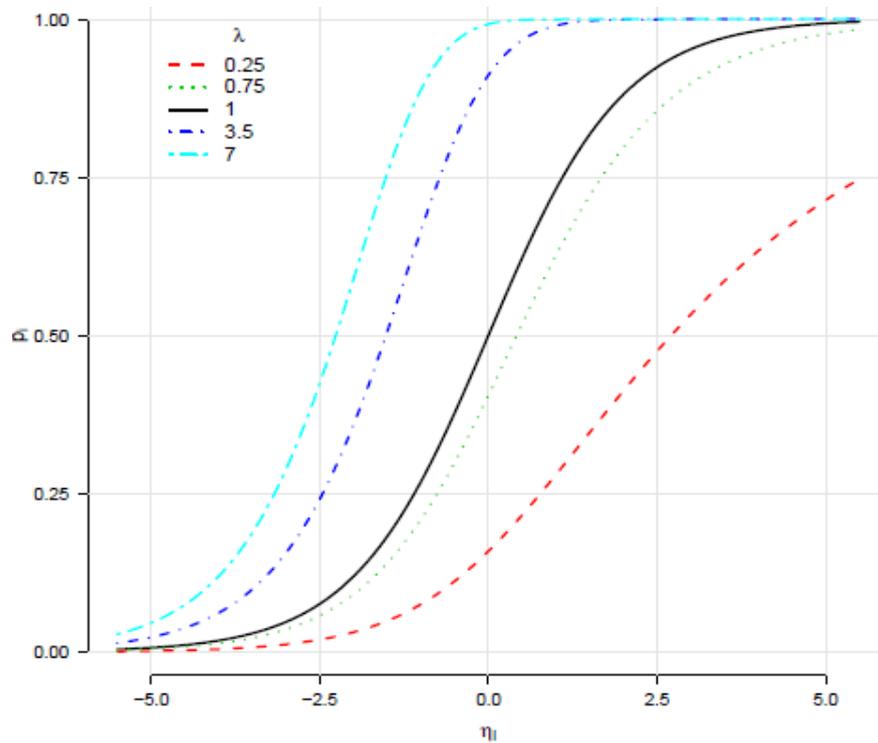
**Tabela A2:** Funções de ligação assimétrica para regressão binária a partir das distribuições potência e reversa de potência obtidas das distribuições básicas.

Nome da ligação	Distribuição	$F_{\lambda}(\cdot)$	$F_{\lambda}^{-1}(\cdot)$
Potência Logit (PL)	PL	$p = \left( \frac{\exp(\eta)}{1 + \exp(\eta)} \right)^{\lambda}$	$\eta = \log \left( \frac{\frac{1}{1-p^{\lambda}}}{1-p^{\lambda}} \right)$
Potência Probit (PP)	PN	$p = (\phi(\eta))^{\lambda}$	$\eta = \phi^{-1}(p^{\lambda})$
Potência Cauchit (PC)	PC	$p = \left( 0,5 + \frac{\arctan(\eta)}{\pi} \right)^{\lambda}$	$\eta = \tan(\pi(p^{\frac{1}{\lambda}} - 0,5))$
Potência Cloglog (PCLL)	PGVMi	$p = (1 - \exp(-\exp(\eta)))^{\lambda}$	$\eta = \log(-\log(1 - p^{\lambda}))$
Potência loglog (PLL)	PGVMa	$p = (\exp(-\exp(-\eta)))^{\lambda}$	$\eta = -\log(-\log(p^{\lambda}))$
Reversa de Potência logit (RPL)	RPL	$p = 1 - \left( \frac{\exp(-\eta)}{1 + \exp(-\eta)} \right)^{\lambda}$	$\eta = \log \left( \frac{1 - (1-p)^{\lambda}}{(1-p)^{\lambda}} \right)$
Reversa de Potência Probit (RPP)	RPP	$p = 1 - (\phi(-\eta))^{\lambda}$	$\eta = -\phi^{-1}((1-p)^{\lambda})$
Reversa de Potência Cauchit (RPC)	RPC	$p = 1 - \left( 0,5 + \frac{\arctan(-\eta)}{\pi} \right)^{\lambda}$	$\eta = -\tan(\pi((1-p)^{\frac{1}{\lambda}} - 0,5))$
Reversa de Potência Cloglog (RPCLL)	RPCLL	$p = 1 - (1 - \exp(-\exp(-\eta)))^{\lambda}$	$\eta = -\log(-\log(1 - (1-p)^{\lambda}))$
Reversa de Potência Loglog (RPLL)	RPLL	$p = 1 - (\exp(-\exp(\eta)))^{\lambda}$	$\eta = \log(-\log((1-p)^{\lambda}))$

Fonte: Adaptado de Anyosa (2017).

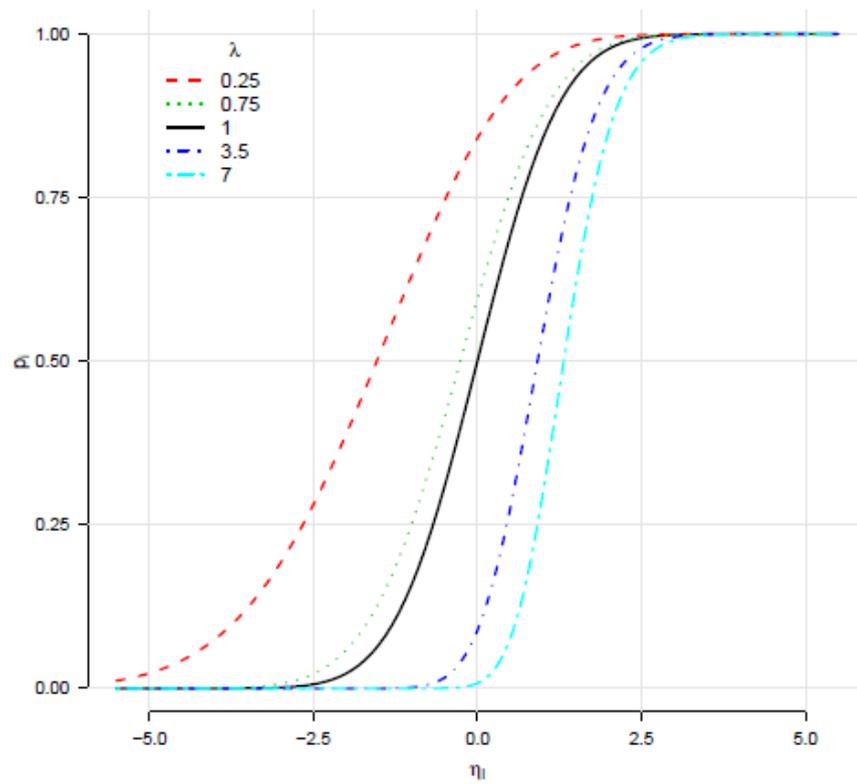
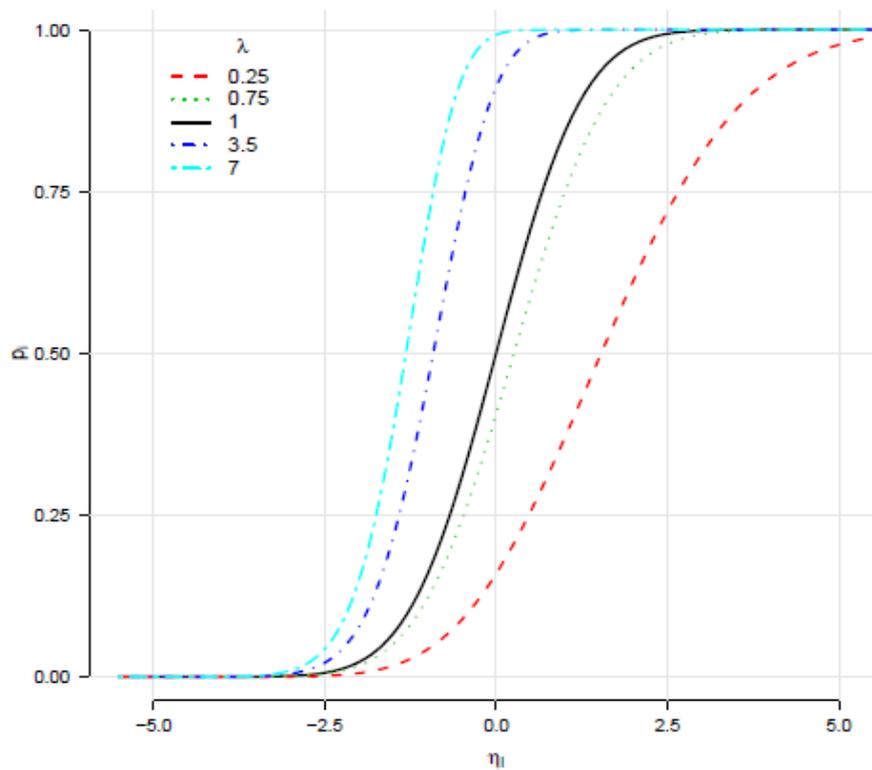
## Interpretação do parâmetro de Assimetria

Pode-se verificar como o parâmetro  $\lambda$  influencia na assimetria da f.d.a. das distribuições potência e reversa de potência e, conseqüentemente, influencia a assimetria nas funções de ligação obtidas a partir destas distribuições. Nas Figuras A1 a A5, são apresentadas as curvas teóricas de probabilidade de sucesso considerando as ligações potência e reversa de potência para alguns valores de  $\lambda = (0,25; 0,75; 1; 3; 5; 7)$ . Nestas figuras, a linha contínua representante da curva de resposta da distribuição base, isto é,  $\lambda = 1$ . Para gerar tais curvas, supomos que  $\eta = \beta_0 + \beta_1 x_{i1}$ , com  $x_{i1} \in [-a, a]$ , com  $a = 5$ ,  $\beta_0 = 0$  e  $\beta_1 = 1$ .

(a) Função de ligação *PL*.(b) Função de ligação *RPL*.

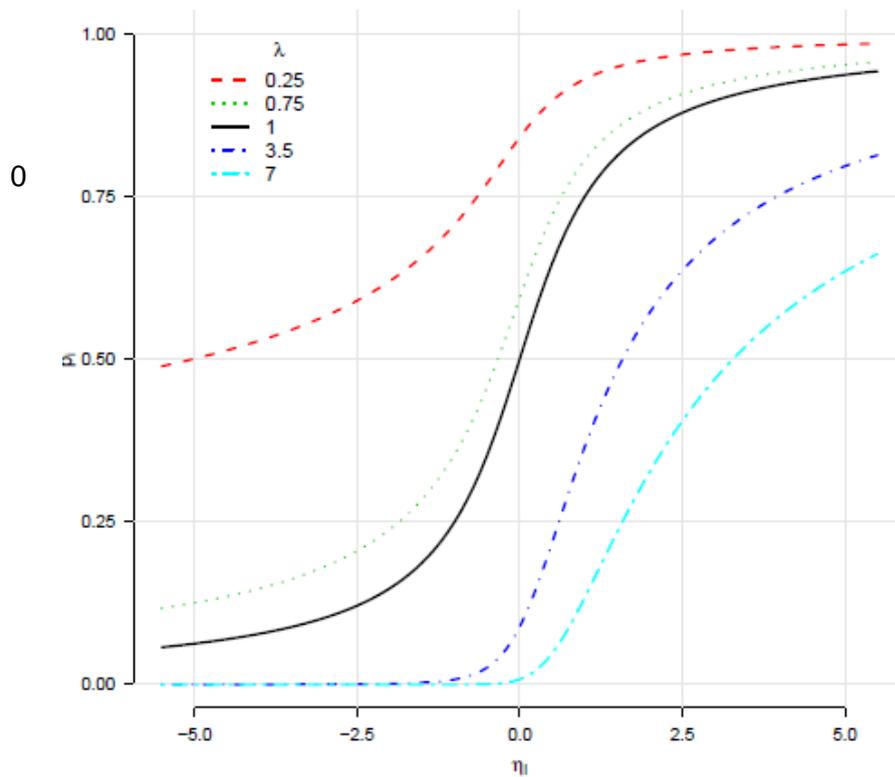
**Figura A1:** Curvas de probabilidade de sucesso considerando as funções de ligação *PL* (a) e *RPL* (b).

Fonte: Galo (2020).

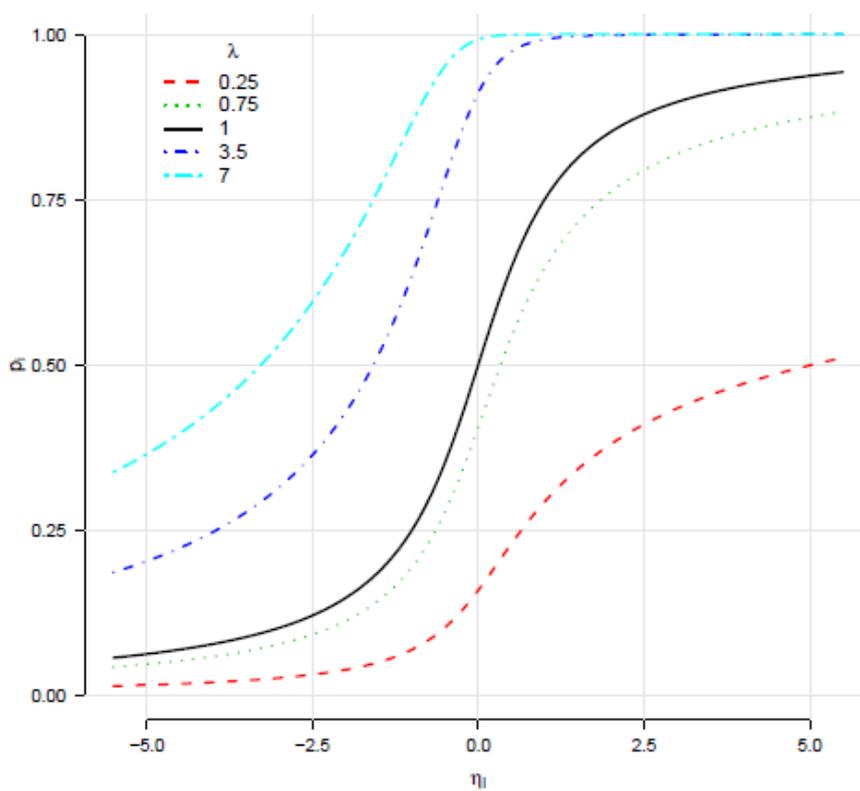
(a) Função de ligação *PP*.(b) Função de ligação *RPP*.

**Figura A2:** Curvas de probabilidade de sucesso considerando as funções de ligação *PP* (a) e *RPP* (b).

Fonte: Galo (2020).



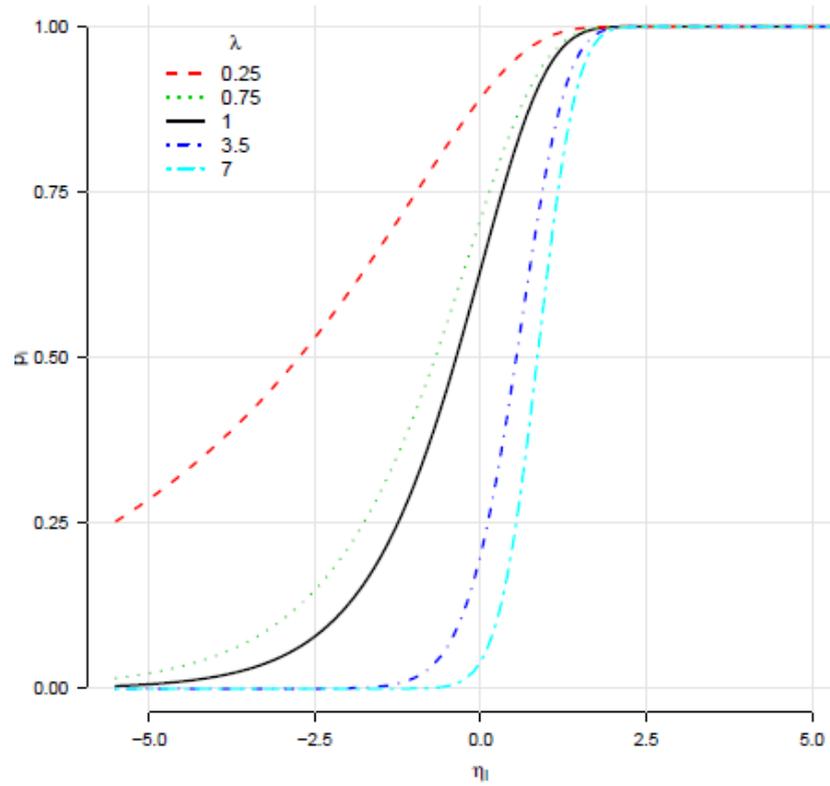
(a) Função de ligação PC.



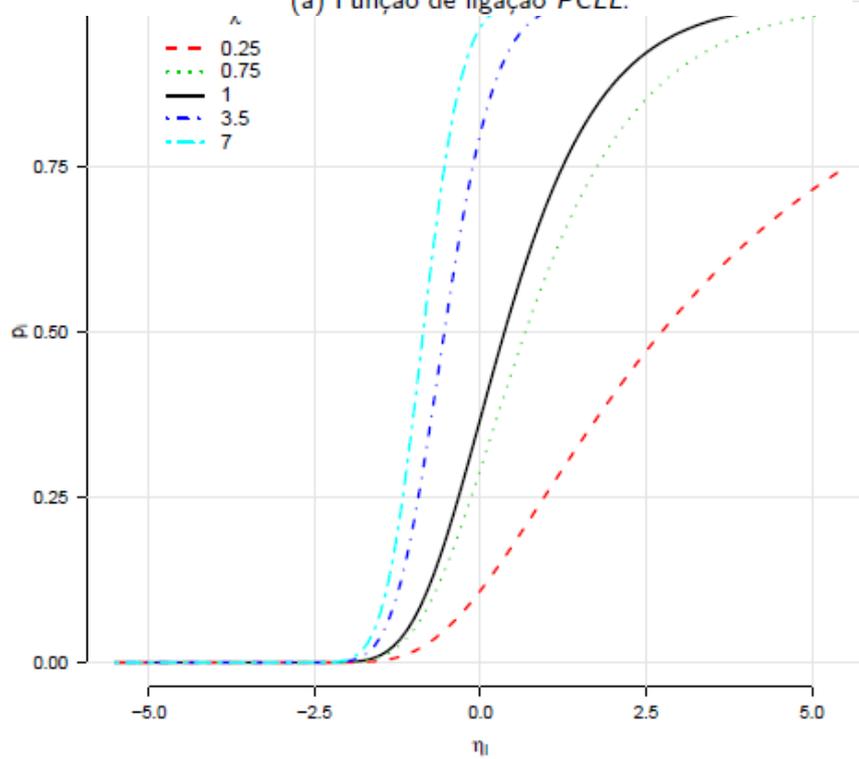
(b) Função de ligação RPC.

**Figura A3:** Curvas de probabilidade de sucesso considerando as funções de ligação PC (a) e RPC (b).

Fonte: Galo (2020).



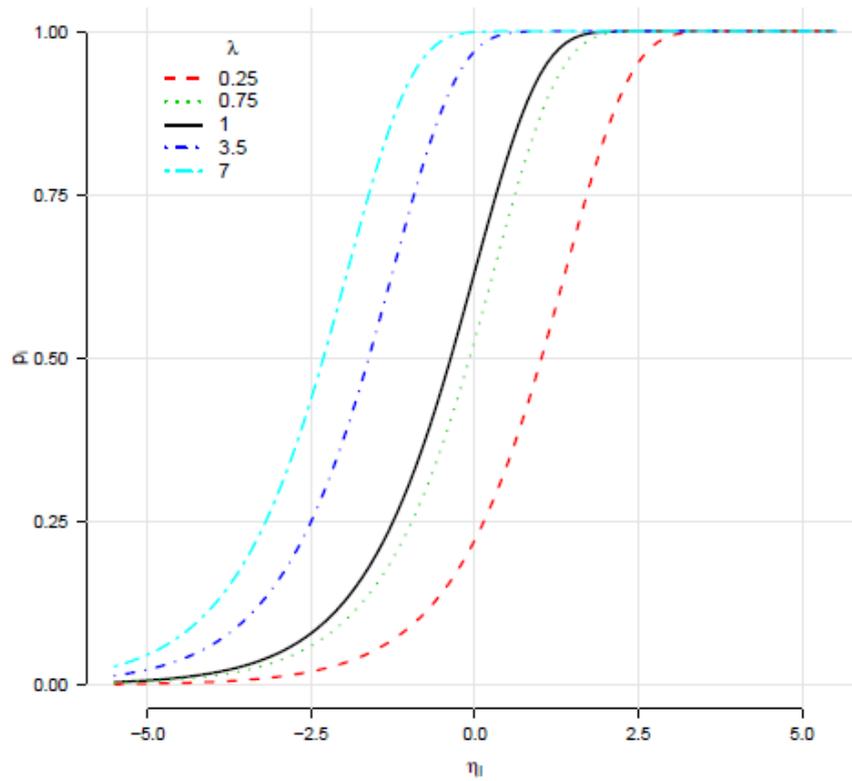
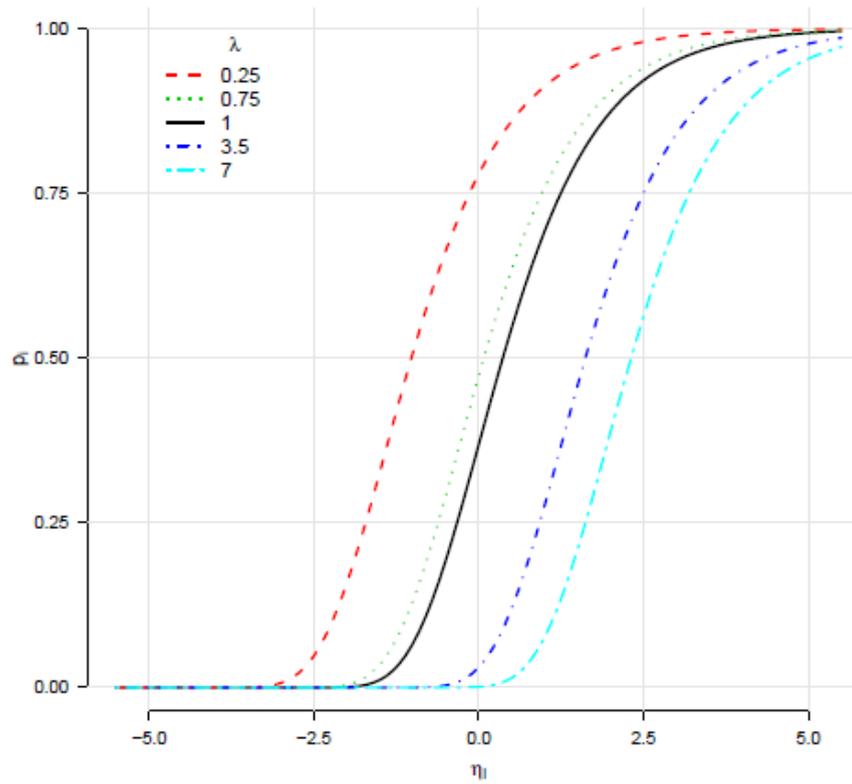
(a) Função de ligação PCLL.



(b) Função de ligação RPCLL.

**Figura A4:** Curvas de probabilidade de sucesso considerando as funções de ligação PCLL (a) e RPCLL (b).

Fonte: Galo (2020).



(b) Função de ligação *RPLL*.

**Figura A5:** Curvas de probabilidade de sucesso considerando as funções de ligação PLL (a) e RPLL (b).

Fonte: Galo (2020).

Anyosa (2017) conclui em seus estudos de simulação que, de maneira geral o valor que  $\lambda$  assume, sofre uma forte influência da proporção de sucessos na variável resposta. No caso das ligações potências, quando  $0 < \lambda < 1$  verifica-se uma maior proporção de sucessos do que falhas na variável resposta; para  $\lambda = 1$  a proporção de sucessos e falhas estão próximas ao equilíbrio e para  $\lambda > 1$  a proporção de sucessos é menor do que a de falhas. Os resultados inversos são verificados para as ligações reversa de potência.

De maneira mais simples, a proporção de sucesso ou fracasso na resposta pode ser vista a partir do deslocamento que ocorre nas curvas de probabilidade. Quando  $\lambda = 1$ , a curva é a mesma da distribuição base utilizada para determinar a distribuição potência ou reversa de potência; para  $0 < \lambda < 1$ , as curvas para as ligações potência apresentam deslocamento à esquerda enquanto as curvas para as ligações reversas de potência apresentam deslocamento à direita da curva da ligação básica; já para  $\lambda > 1$ , as curvas das ligações potência apresentam deslocamento à direita, enquanto as curvas das ligações reversas de potência apresentam deslocamento à esquerda da curva da ligação básica.

Para os modelos descritos serão aplicadas as funções de ligação *logit*, *probit*, *cauchit*, *loglog*, e *cloglog*, além de suas respectivas variações de potência e reversa de potência. A escolha de múltiplas funções visa não só explorar qual delas melhor se adapta aos dados de maturação do curimba (*Prochilodus lineatus*), mas também permitir uma análise comparativa entre as alternativas, considerando tanto a precisão dos ajustes quanto o tempo de processamento.

## APÊNDICE B

### ESTIMATIVAS A POSTERIORI DOS PARÂMETROS COM SEUS RESPECTIVOS INTERVALOS HPD

**Tabela B1:** Estimativas (médias) *a posteriori* dos parâmetros com seus respectivos intervalos HPD dos modelos com funções de ligação básicas apresentadas na tabela 1.

Função de ligação	Parâmetros	Média	Mediana	Desvio padrão	HPD
L	$\beta_0$	-15,058	-15,06	0,298	(-15,257 ; -14,852)
	$\beta_1$	0,585	0,585	0,013	(0,576 ; 0,594)
	$\beta_2$	0,051	0,051	0,003	(0,049 ; 0,054)
	$L50f$	25,725	25,724	0,124	(25,647 ; 25,811)
	$L50m$	23,654	23,65	0,089	(23,596 ; 23,713)
	$L50f - L50m$	2,071	2,07	0,139	(1,982 ; 2,169)
	$L99f$	33,579	33,579	0,267	(33,404 ; 33,765)
	$L99m$	30,876	30,868	0,217	(30,729 ; 31,022)
	$L99f - L99m$	2,704	2,7	0,183	(2,583 ; 2,83)
P	$\beta_0$	-7,55	-7,551	0,134	(-7,64 ; -7,456)
	$\beta_1$	0,29	0,29	0,006	(0,285 ; 0,294)
	$\beta_2$	0,026	0,026	0,002	(0,024 ; 0,027)
	$L50f$	28,465	28,464	0,173	(28,35 ; 28,576)
	$L50m$	26,149	26,145	0,129	(26,062 ; 26,238)
	$L50f - L50m$	2,316	2,315	0,164	(2,209 ; 2,424)
	$L99f$	28,975	28,973	0,181	(28,852 ; 29,091)
	$L99m$	26,617	26,613	0,137	(26,524 ; 26,712)
	$L99f - L99m$	2,357	2,357	0,167	(2,249 ; 2,468)
C	$\beta_0$	-30,877	-31,616	1,127	(-31,597 ; -30,051)
	$\beta_1$	1,224	1,223	0,046	(1,189 ; 1,254)
	$\beta_2$	0,105	1,223	0,007	(0,1 ; 0,110)
	$L50f$	25,225	25,222	0,111	(25,15 ; 25,301)
	$L50m$	23,234	23,231	0,066	(23,187 ; 23,278)
	$L50f - L50m$	1,99	1,989	0,126	(1,906 ; 2,076)
	$L99f$	51,255	51,249	1,027	(50,557 ; 52,014)
	$L99m$	47,21	47,194	0,912	(46,6 ; 47,868)
	$L99f - L99m$	4,045	4,037	0,274	(3,864 ; 4,231)
CLL	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX
LL	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX

**Tabela B2:** Estimativas (médias) *a posteriori* dos parâmetros com seus respectivos intervalos HPD das funções de potência apresentadas na tabela 2.

Função de ligação	Parâmetros	Média	Mediana	Desvio padrão	HPD
PL	$\beta_0$	-13,405	-13,366	0,755	(-13,97 ; -12,955)
	$\beta_1$	0,531	0,53	0,026	(0,516 ; 0,551)
	$\beta_2$	0,046	0,046	0,004	(0,044 ; 0,049)
	$\lambda$	1,209	1,205	0,101	(1,132 ; 1,265)
	$L50f$	25,705	25,706	0,123	(25,624 ; 25,79)
	$L50m$	23,648	23,646	0,09	(23,585 ; 23,705)
	$L50f - L50m$	2,057	2,053	0,137	(1,968 ; 2,156)
	$L99f$	34,253	34,245	0,423	(33,941 ; 34,504)
	$L99m$	31,513	31,508	0,377	(31,224 ; 31,732)
	$L99f - L99m$	2,741	2,737	0,186	(1,968 ; 2,156)
PP	$\beta_0$	-9,66	-9,592	1,297	(-10,487 ; -8,772)
	$\beta_1$	0,351	0,349	0,038	(0,324 ; 0,375)
	$\beta_2$	0,031	0,031	0,004	(0,028 ; 0,034)
	$\lambda$	0,624	0,596	0,185	(0,492 ; 0,721)
	$L50f$	29,446	29,483	0,552	(28,979 ; 29,576)
	$L50m$	27,04	27,067	0,488	(26,624 ; 27,147)
	$L50f - L50m$	2,406	2,4	0,18	(2,284 ; 2,511)
	$L99f$	29,871	29,905	0,512	(29,562 ; 30,222)
	$L99m$	27,431	27,454	0,45	(27,158 ; 27,741)
	$L99f - L99m$	2,44	2,436	0,182	(2,331 ; 2,564)
PC	$\beta_0$	-12,847	-12,819	0,548	(-13,175 ; -12,445)
	$\beta_1$	0,551	0,551	0,022	(0,535 ; 0,565)
	$\beta_2$	0,043	0,043	0,003	(0,041 ; 0,045)
	$\lambda$	2,093	2,091	0,069	(2,047 ; 2,142)
	$L50f$	24,786	24,786	0,111	(24,709 ; 24,862)
	$L50m$	22,973	22,97	0,086	(22,92 ; 23,034)
	$L50f - L50m$	1,812	1,813	0,119	(1,728 ; 1,89)
	$L99f$	144,087	143,938	7,777	(139,37 ; 149,644)
	$L99m$	133,548	133,362	7,168	(129,045 ; 138,726)
	$L99f - L99m$	10,538	10,52	0,915	(9,901 ; 11,144)
PCLL	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX
PLL	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX

**Tabela B3:** Estimativas (médias) *a posteriori* dos parâmetros com seus respectivos intervalos HPD das funções reversas de potência apresentadas na tabela 2.

Função de ligação	Parâmetros	Média	Mediana	Desvio padrão	HPD
RPL	$\beta_0$	-16,299	-16,294	0,463	(-16,625 ; -16)
	$\beta_1$	0,671	0,671	0,025	(0,655 ; 0,689)
	$\beta_2$	0,057	0,057	0,004	(0,055 ; 0,057)
	$\lambda$	0,572	0,567	0,062	(0,525 ; 0,613)
	$L50f$	24,351	24,346	0,265	(24,147 ; 24,528)
	$L50m$	22,431	22,425	0,229	(22,256 ; 22,585)
	$L50f - L50m$	1,919	1,917	0,129	(1,827 ; 2,007)
	$L99f$	28,229	28,181	0,817	(27,597 ; 28,753)
	$L99m$	26,003	25,96	0,74	(25,433 ; 26,468)
	$L99f - L99m$	2,225	2,219	0,164	(2,104 ; 2,334)
RPP	$\beta_0$	-6,813	-6,809	0,176	(-6,939 ; -6,695)
	$\beta_1$	0,229	0,229	0,013	(0,22 ; 0,238)
	$\beta_2$	0,021	0,020	0,002	(0,019 ; 0,022)
	$\lambda$	3,281	3,107	0,991	(2,577 ; 3,819)
	$L50f$	26,383	26,335	0,787	(25,83 ; 26,872)
	$L50m$	24,211	24,166	0,715	(23,706 ; 24,667)
	$L50f - L50m$	2,171	2,171	0,167	(2,06 ; 2,283)
	$L99f$	27,226	27,203	0,74	(26,715 ; 27,7)
	$L99m$	24,985	24,963	0,67	(24,521 ; 25,421)
	$L99f - L99m$	2,241	2,241	0,17	(2,128 ; 2,238)
RPC	$\beta_0$	-39,283	-39,228	2,154	(-40,611 ; -39,181)
	$\beta_1$	1,695	1,631	0,095	(1,63 ; 1,754)
	$\beta_2$	0,127	0,126	0,011	(0,119 ; 0,134)
	$\lambda$	0,393	0,392	0,016	(0,381 ; 0,403)
	$L50f$	24,17	24,169	0,097	(24,105 ; 24,236)
	$L50m$	22,491	22,488	0,088	(22,434 ; 22,548)
	$L50f - L50m$	1,679	1,679	0,11	(1,609 ; 1,752)
	$L99f$	26476,883	23394,475	13588,819	(17580,366 ; 32582,408)
	$L99m$	24631,09	21731,231	12625,618	(16366,176 ; 30365,228)
	$L99f - L99m$	1845,793	1615,661	972,139	(1197,703 ; 2276,408)
RPCLL	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX
RPLL	XXXXXX	XXXXXX	XXXXXX	XXXXXX	XXXXXX

## APÊNDICE C

---

### CÓDIGOS UTILIZADOS PARA A MODELAGEM

---

#### Processo de estimação

# Analises Bayesianas, modelos dicotômicos.

```
dadosb = cbind(X = db$LS,n = 1,y = db$REP, z = db$SEXO)
```

```
dadosb = na.omit(dadosb)
```

```
y = dadosb[,"y"]
```

```
X = dadosb[,'X']
```

```
Z = dadosb[,'z']
```

```
K = NCOL(X)
```

```
dados.aux <- list(K=K, X=X, n=length(y), y=y, z=Z, p0=0.5, p1=0.99)
```

```
dadosb2 = aggregate(cbind(y,n) ~ X + z, data = dadosb, FUN = sum)
```

```
dados.aux2 <- list(X=dadosb2$X, N=NROW(dadosb2), y=dadosb2$y,Z = dadosb2$z,n =  
dadosb2$n, p0=0.5, p1=0.99, pi = pi)
```

```
iter=300000
```

```
burn=100000
```

```
salto=400
```

```
nchains = 6
```

```
alfa <- 0.05
```

```
parametros <- c("L100m","L100f","Diff1","L50m", "L50f", "Diff", "beta0", "beta1", "beta2",  
"delta", "lambda")
```

```
#-----#
```

```
#familia basica
```

```
# Funcao Bayesiana para Regressao Binomial
```

```
# Funcao de Ligacao logit
```

```
# nu = beta0 + beta*X
```

```
# p = exp(nu)/(1+exp(nu))
```

```
# y: numero de sucessos
```

```
# X: matriz de incidencia ou (co)variaveis
```

```

sink("logit.txt")
cat("
  model {
    for (i in 1:N) {
      y[i] ~ dbinom(p[i], n[i])
      nu[i] <- beta0 + (beta1 + beta2*Z[i]) * X[i]
      pl[i] <- (exp(nu[i])) / (1 + exp(nu[i]))
      p[i] <- pow(pl[i], lambda)
    }
    # Distribuições a priori
    beta0 ~ dnorm(0, 0.000001)
    beta1 ~ dnorm(0, 0.000001)
    beta2 ~ dnorm(0, 0.000001)
    #delta ~ dunif(-2, 2)
    lambda <- 1
    # L50
    L50m <- (log(p0^(1/lambda) / (1 - p0^(1/lambda)))) - beta0 / (beta1 + beta2)
    L50f <- (log(p0^(1/lambda) / (1 - p0^(1/lambda)))) - beta0 / beta1
    Diff <- L50f-L50m
    #L100
    L100m <- (log(p1^(1/lambda) / (1 - p1^(1/lambda)))) - beta0 / (beta1 + beta2)
    L100f <- (log(p1^(1/lambda) / (1 - p1^(1/lambda)))) - beta0 / beta1
    Diff1 <- L100f-L100m
  }
", fill=TRUE)
sink()
res.freq <- with(dadosb2, glm(cbind(y,n-y) ~ X + X:z, family=binomial(link="logit")))
beta.aux <- coef(res.freq)
chutes.aux <- function() list(beta0=beta.aux[1], beta1=beta.aux[2], beta2=beta.aux[3],
delta=runif(1, -2, 2))
system.time(fit_L <- jags.parallel(model.file='logit.txt', data=dados.aux2, inits=chutes.aux,
n.chains=nchains, n.iter = iter, n.burnin = burn, n.thin = salto, parameters.to.save =
parametros,
export_obj_names=c('iter','burn','salto', 'beta.aux'))
)

```

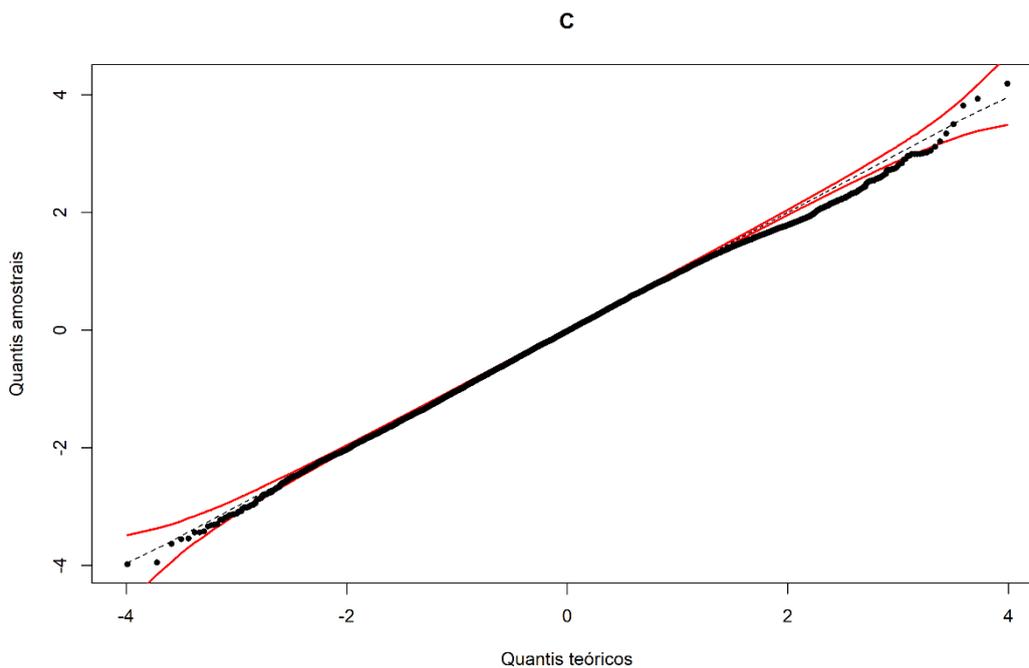
```
traplot(fit_L)
denplot(fit_L)
L_cadeias <- fit_L$BUGSoutput$sims.matrix
print(fit_L)
HPDinterval(mcmc(L_cadeias), prob=0.95)
print(geweke.diag(L_cadeias))
print(heidel.diag(L_cadeias))
#-----#
```

## APÊNDICE D

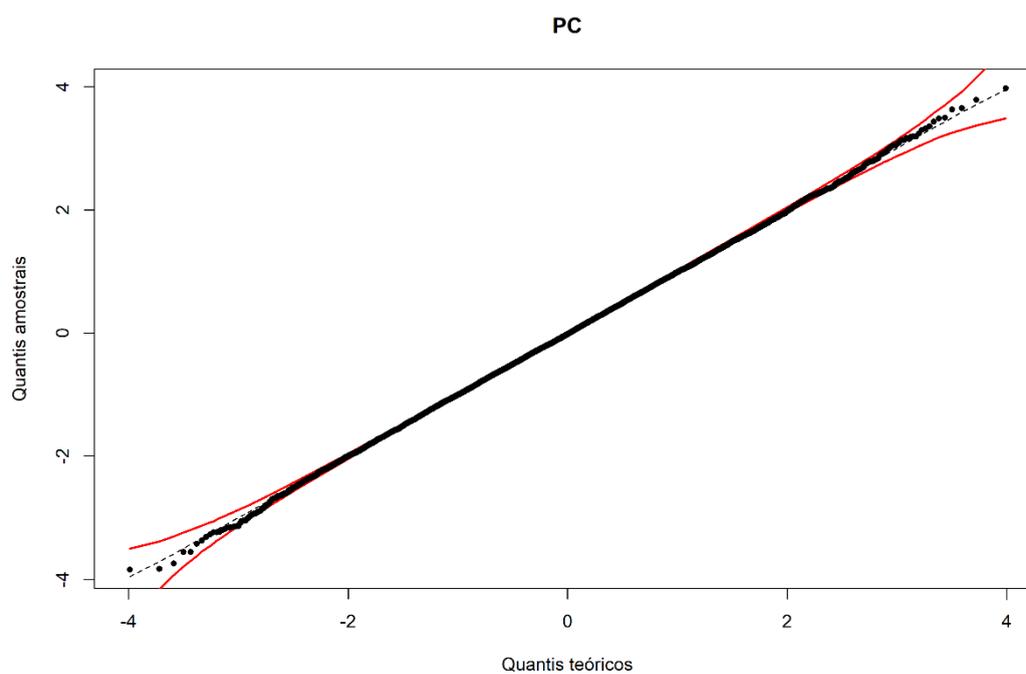
## CÓDIGOS UTILIZADOS PARA A MODELAGEM

## Envelopes simulados.

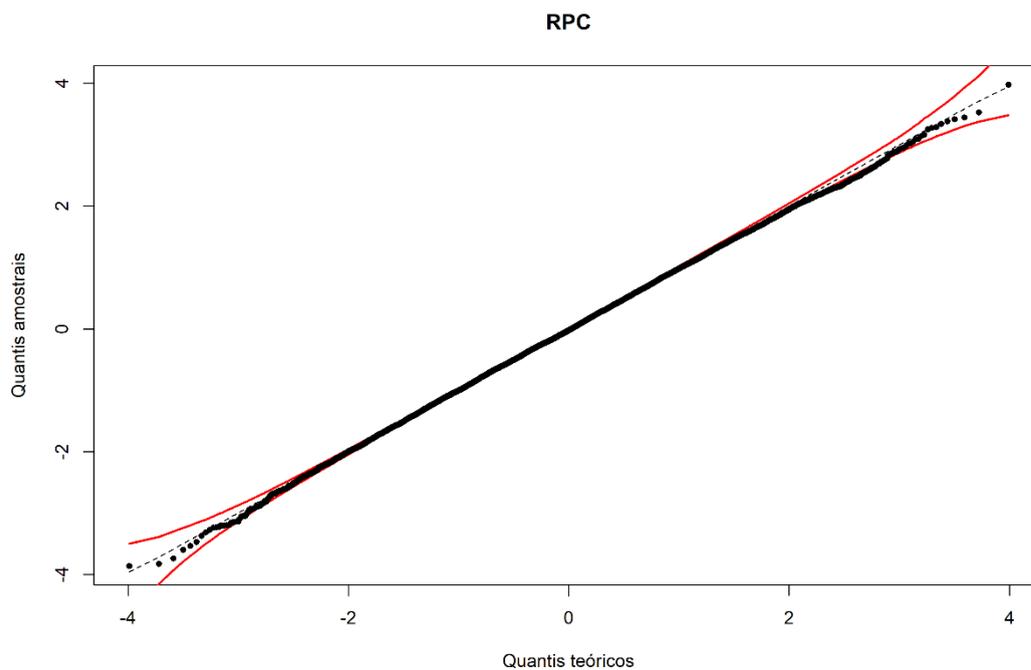
Da Figura D1 à Figura D9 são apresentados os envelopes simulados de cada modelo ajustado.



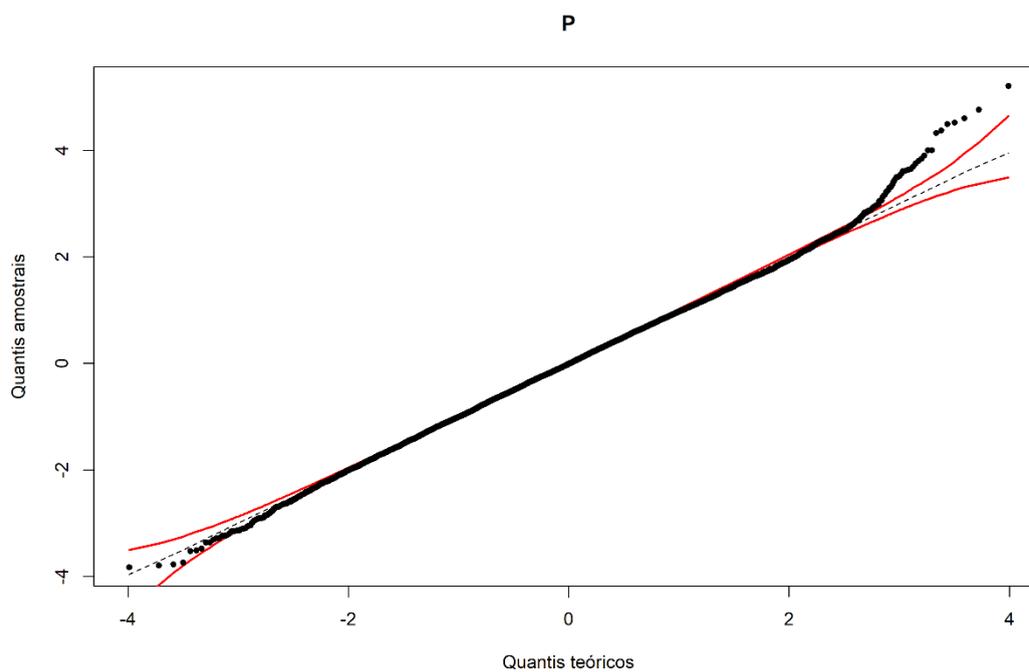
**FIGURA D1:** Envelope simulado para o modelo *Cauchit*.



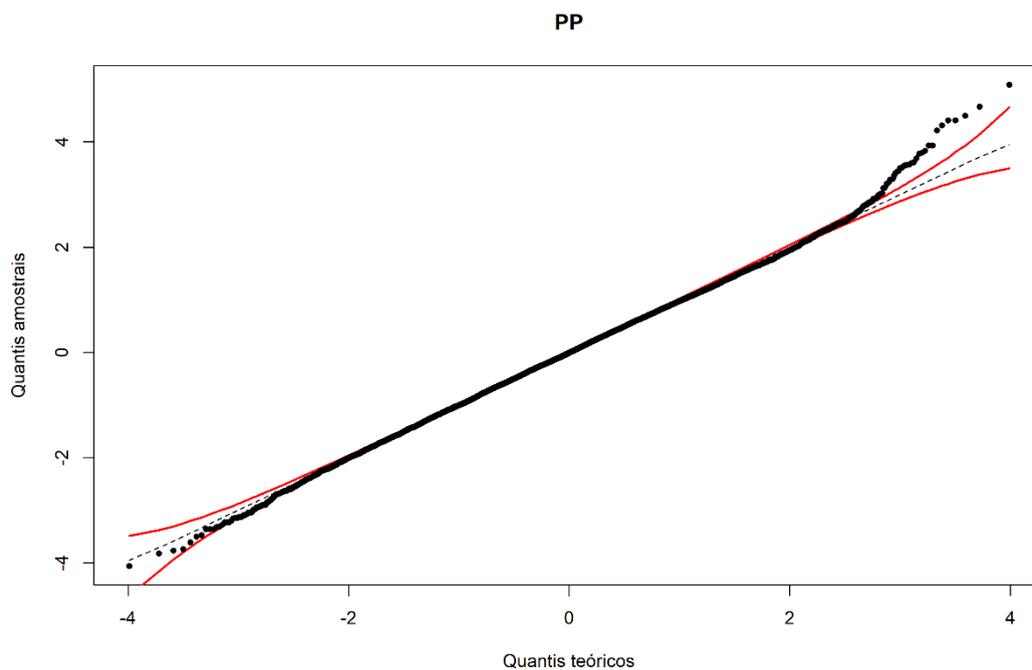
**FIGURA D2:** Envelope simulado para o modelo *Potencia Cauchit*.



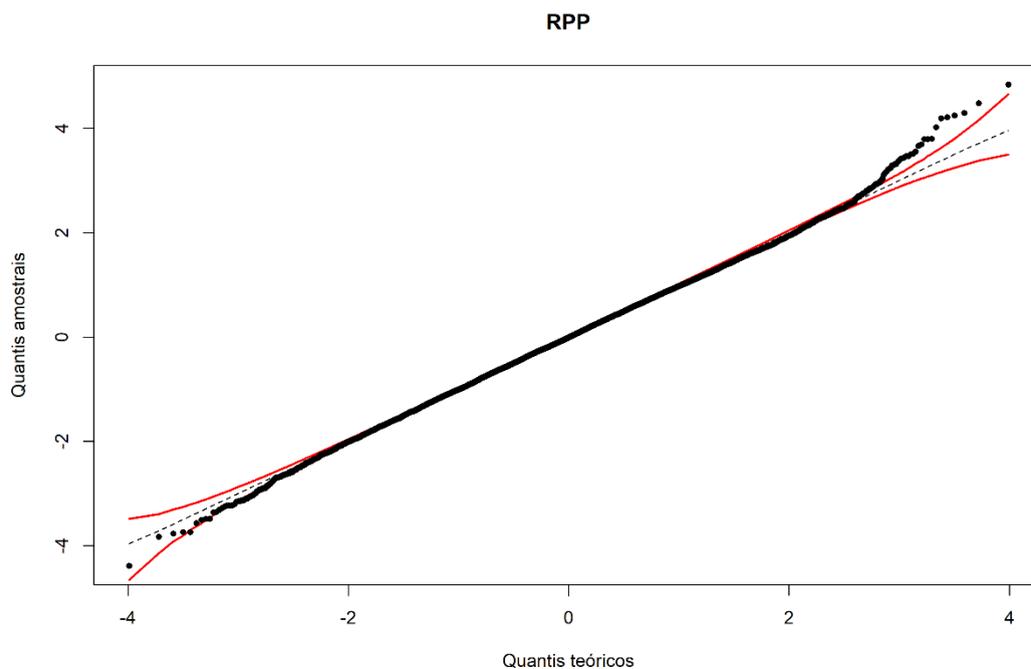
**FIGURA D3:** Envelope simulado para o modelo *Reversa de Potencia Cauchit*.



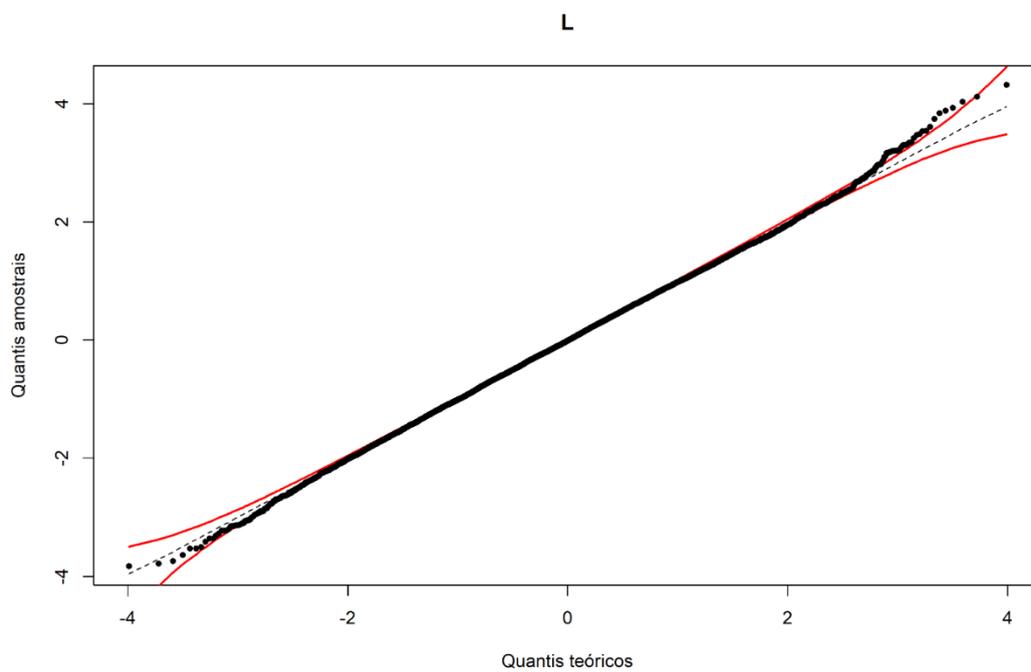
**FIGURA D4:** Envelope simulado para o modelo *Probit*.



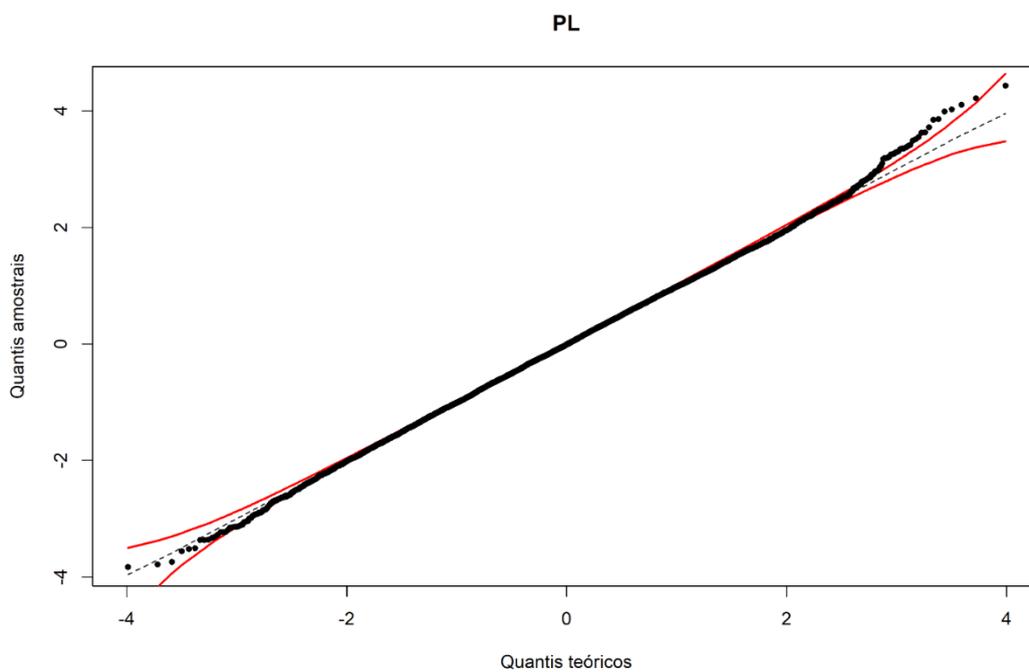
**FIGURA D5:** Envelope simulado para o modelo *Potência Probit*.



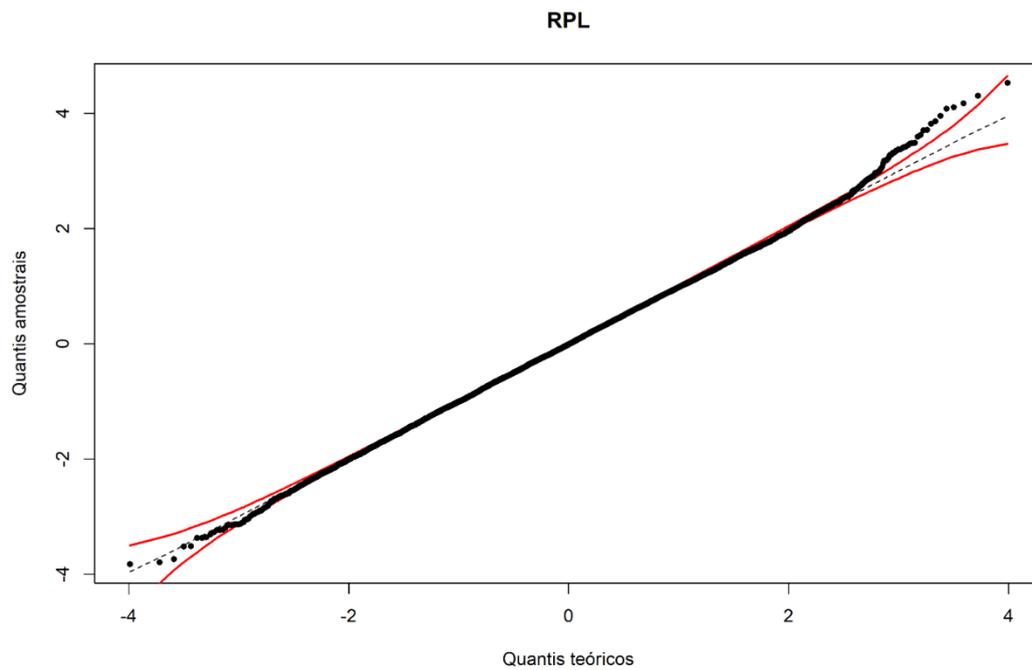
**FIGURA D6:** Envelope simulado para o modelo *Reversa Potência Probit*.



**FIGURA D7:** Envelope simulado para o modelo *Logit*.



**FIGURA D8:** Envelope simulado para o modelo *Potência Logit*.



**FIGURA D9:** Envelope simulado para o modelo *Reversa de Potência Logit*.