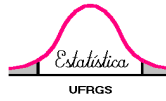




UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA



# Introdução à Regressão Logística Quantílica

Autor: Tiago Buchweitz Correa  
Orientador: Professor Dr. Álvaro Vigo

Porto Alegre, 11 de dezembro de 2015.

Universidade Federal do Rio Grande do Sul  
Instituto de Matemática e Estatística  
Departamento de Estatística

# Introdução à Regressão Logística Quantílica

Autor: Tiago Buchweitz Correa

Trabalho de Conclusão de Curso  
apresentado para obtenção  
do grau de Bacharel em Estatística.

Banca Examinadora:  
Professor Dr. Álvaro Vigo  
Professora Dra. Luciana Neves Nunes

Porto Alegre, 11 de dezembro de 2015.

*Aos meus pais, Cesar Augusto Cunha Correa e Christine Buchweitz  
Correa.*

## **Agradecimentos**

Agradeço meu pai, Cesar, que dentre todas coisas boas, me deu todas as condições e todo o apoio para que concluísse a faculdade. É o meu maior modelo e queria muito que pudesse estar comigo nesse momento. Tenho certeza que está muito feliz onde quer que esteja.

Agradeço minha mãe, Christine, por nunca desistir e me dar forças por todo esse tempo.

Agradeço a minha namorada, Lara, por me fazer continuar e ajudar na conclusão do trabalho.

Por fim, agradeço, e muito, ao meu orientador, Álvaro, por todo seu tempo investido em mim.

Tenho certeza que serei eternamente grato a todos por esse momento.

## Resumo

Este trabalho apresenta a Regressão Logística Quantílica como uma alternativa para analisar respostas com desfechos limitados e estimar associações e preditores de interesse. O modelo foi proposto por Bottai, Cai e McKeown (2010, *Statistics in Medicine* 29: 309-317) e tem como premissa modelar os quantis da resposta, sendo vantajoso para análise exploratória de associações no contexto em que existe grande variabilidade. As interpretações são análogas às do modelo de regressão logística tradicional. Através de transformações definidas pelo modelo e com ajuda do software SAS e procedimentos computacionais para regressão quantílica, foram analisados os quantis de 5, 25, 50, 75 e 95% de um exemplo, onde é possível ilustrar facilmente os resultados e identificar as razões de chance da variável resposta em relação aos preditores.

**Palavras-chave:** regressão logística, regressão quantílica, resposta limitada, odds ratio.

## Sumário

<b>1 INTRODUÇÃO .....</b>	<b>7</b>
<b>2 OBJETIVO .....</b>	<b>8</b>
<b>3 REVISÃO DA LITERATURA .....</b>	<b>9</b>
<b>4 EXEMPLO .....</b>	<b>12</b>
<b>4.1 Geração dos dados.....</b>	<b>12</b>
<b>4.2 Ajuste do modelo.....</b>	<b>16</b>
<b>5 CONSIDERAÇÕES FINAIS .....</b>	<b>27</b>
<b>6 REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>28</b>
<b>7 ANEXO – ROTINA SAS PARA AJUSTE DOS MODELOS.....</b>	<b>29</b>

## 1 INTRODUÇÃO

Este trabalho apresenta o modelo de regressão logística quantílica proposto por Bottai, Cai e McKeown (2010) como uma alternativa para analisar a variabilidade de um desfecho que assume valores em um intervalo limitado, condicional a um conjunto de covariáveis. O modelo é baseado em uma regressão quantílica com transformação logística dos quantis da variável com intervalo limitado e é válido para qualquer distribuição subjacente. Importante salientar que existem poucos trabalhos sobre este assunto e todos utilizam este mesmo artigo de Bottai, Cai e McKeown (2010) como referência.

A abordagem pode ser útil no contexto em que a variabilidade da resposta é grande ou, ainda, quando as relações com os preditores são não lineares. É possível fazer inferências sobre associações de diferentes quantis da distribuição da resposta e não apenas sobre a média. O método é robusto a outliers e potencialmente vantajoso principalmente quando a distribuição condicional do desfecho é assimétrica ou possui variância heterogênea.

Por ser um modelo ainda pouco explorado, sua metodologia é explicada com objetivo de criar um entendimento mais amplo de sua aplicação. A descrição de modelagem de dados e rotinas computacionais são exemplificadas para que seja possível que o leitor possa tirar suas próprias conclusões e identificar as possíveis vantagens propostas pelos autores do modelo no exemplo criado.

## **2 OBJETIVO**

Apresentar os conceitos básicos do método de regressão logística quantílica para desfechos quantitativos limitados, bem como ilustrar como pode ser implementado utilizando procedimentos usuais de regressão quantílica.



### 3 REVISÃO DA LITERATURA

Variáveis respostas que assumem valores em um intervalo delimitado ocorrem com frequência em estudos científicos ou em aplicações de diversas áreas. Exemplos conhecidos são instrumentos para avaliar doenças mentais, cuja intensidade de sintomas é medida por meio de escores baseados na Teoria Clássica dos Testes<sup>1</sup>, como DSM IV<sup>2</sup>, CES-D<sup>3</sup> e Mini Mental<sup>4</sup>. Também, pode-se citar escores criados pela soma das respostas aos itens de um instrumento de pesquisa, usando, por exemplo, a escala de Likert<sup>5</sup> como pesquisas sobre satisfação de clientes.

Nestes contextos podem ocorrer grande variabilidade nos escores das respostas, bem como assimetria em suas distribuições. Tais características podem dificultar o ajuste de modelos para relacionar o escore com preditores de interesse. Como alternativas de análise, surgem o uso de modelos de regressão quantílica ou de regressão logística quantílica. Ao contrário do primeiro, que modela os quantis da variável de resposta, condicional nas covariáveis, o segundo consiste em modelar a probabilidade de assumir valor da resposta igual ou superior a um determinado quantil. Ambas abordagens permitem identificar diferenças nas associações entre os quantis da resposta e os preditores, o que se torna uma ferramenta mais poderosa para fazer comparações, não só pela média, mas também por outros aspectos, além de permitir estimar as razões de chances para cada quantil modelado a partir das estimativas dos coeficientes de regressão.

Considere uma variável resposta quantitativa  $y$  e um conjunto de preditores  $x = (x_1, x_2, \dots, x_s)$ . Para uma amostra de  $n$  observações  $y_i$ ,  $i = 1, \dots, n$

---

<sup>1</sup> Escore baseado na soma dos pontos das respostas.

<sup>2</sup> Manual Diagnóstico e Estatístico de Transtornos Mentais – manual para profissionais da área de saúde mental que lista diferentes categorias de transtornos mentais e critérios para diagnosticá-los, de acordo com a Associação Americana de Psiquiatria.

<sup>3</sup> Medida de severidade de depressão.

<sup>4</sup> Exame de estado mental.

<sup>5</sup> Escala da soma das respostas dadas em questionários.

e  $x_i = \{x_{1,i}, \dots, x_{s,i}\}^T$ . Segundo Koenker (2005), o modelo de regressão quantílica especifica que

$$y_i = x_i^T \beta_p + \varepsilon_i$$

onde  $\beta_p = \{\beta_{p,0}, \beta_{p,1}, \dots, \beta_{p,s}\}^T$  representa os parâmetros desconhecidos da regressão e  $p$  o quantil da distribuição da resposta. Para qualquer valor  $p \in (0,1)$ , assume-se que  $P(\varepsilon_i \leq 0 | x_i) = p$  ou, equivalentemente, que  $P(y_i \leq x_i^T \beta_p | x_i) = p$ . O quantil  $p$  da distribuição condicional  $y_i$  dado  $x_i$  é definido por

$$Q_y(p) = x_i^T \beta_p \tag{1}$$

Se  $p = 0,5$ , então  $Q_y(0,5)$  é a mediana condicional, o valor que divide a distribuição da variável de resposta em duas partes com probabilidades iguais. Nenhuma outra suposição é necessária para a distribuição dos resíduos  $\varepsilon_i$ .

Ao contrário da regressão sobre a média, a estimativa da regressão quantílica é invariante a outras transformações monótonas, isto é,  $Q_{h(y)}(p) = h\{Q_y(p)\}$  para qualquer função não decrescente  $h$ , enquanto  $E\{h(y)\} \neq h\{E(y)\}$ , onde  $E(y)$  é a média de  $y$ . Bottai, Cai e McKeown (2010) exploram essa propriedade e definem o modelo de regressão logística quantílica para modelar variáveis contínuas com intervalos limitados conhecidos como

$$y_i \in (y_{min}, y_{max}) \tag{2}$$

onde  $y_{min}$  e  $y_{max}$  são os valores mínimo e máximo (teóricos, não necessariamente observados) da resposta.

Considerando os limites especificados na equação (2), assume-se que para qualquer quantil  $p$  existe um conjunto fixo de parâmetros  $\beta_p$  e uma função conhecida  $h$  não decrescente do intervalo  $(y_{min}, y_{max})$  para a reta real, chamada de *link* dada por

$$h\{Q_y(p)\} = x_i^T \beta_p = \beta_{0,p} + \beta_{1,p}x_1 + \dots + \beta_{s,p}$$

Tendo em vista que uma variável contínua limitada em um intervalo se assemelha a uma probabilidade, Bottai, Cai e McKeown (2010) optam por usar uma transformação logística para a função *link*  $h$

$$h(y_i) = \log\left(\frac{y_i - y_{min}}{y_{max} - y_i}\right) = \text{logit}(y_i), \quad (3)$$

com transformação inversa dada por

$$Q_y(p) = \frac{\exp(x_i^T \beta_p) y_{max} + y_{min}}{1 + \exp(x_i^T \beta_p)}$$

Os coeficientes da regressão podem ser estimados usando regressão quantílica utilizando uma regressão dos resultados transformados  $h(y_i)$  em  $x$  usando (1)

$$Q_{h(y_i)}(p) = Q_{\text{logit}(y_i)}(p) = x_i^T \beta_p$$

Isso é análogo à regressão logística, porém usando uma transformação para modelar uma probabilidade. Essas transformações têm o mesmo objetivo nos dois modelos, que é facilitar a modelagem enquanto restringe a inferência sobre o resultado dentro de um intervalo  $(0, 1)$  para probabilidade e  $(y_{min}, y_{max})$  para o resultado limitado contínuo.

Mesmo sendo simples, o modelo de regressão logística quantílica não está disponível nos softwares de análises comuns. Porém, procedimentos de regressão quantílica usuais podem ser utilizados visto que os quantis da distribuição da resposta são invariantes com respeito às transformações monótonas, como definido na equação (3), podendo, naturalmente, ser utilizadas outras transformações.

O próximo capítulo utiliza um exemplo hipotético com dados simulados para mostrar passos para o ajuste do modelo e interpretações de resultados.

## 4 EXEMPLO

Este capítulo descreve o ajuste do modelo de regressão logística quantílica considerando uma resposta limitada  $y$  no intervalo entre 0 e 60 com três preditores  $X_1, X_2$  e  $X_3$ .

### 4.1 Geração dos dados

O conjunto de dados foi gerado através do software SAS considerando a matriz de correlação dos preditores  $X_1, X_2$  e  $X_3$

$$\rho = \begin{bmatrix} 1 & 0,6 & -0,4 \\ 0,6 & 1 & -0,7 \\ -0,4 & -0,7 & 1 \end{bmatrix}$$

a partir do qual foram geradas 5.000 observações independentes, seguindo os passos:

- 1) Foi gerada a matriz  $Z$  com dimensão (5000x3), a partir da distribuição normal multivariada com vetor de médias (0,0,0), variâncias (1,1,1) e matriz de correlações  $\rho$ ;
- 2) Foi definida a matriz  $U$  (5000 x 3) por meio da transformação  $U_j(Z_j) = \Phi(Z_j), j = 1,2,3$ , em que  $\Phi(\cdot)$  é a função de distribuição da normal padrão. Desse modo, as colunas de  $U$  são variáveis aleatórias com distribuição uniforme em (0,1), porém não são independentes;
- 3) Foi definida a matriz  $V$  (5000x3), por meio da transformação  $V_j = F^{-1}(U_j)$ , em que  $F^{-1}(\cdot)$  é a inversa da função de distribuição especificada: para  $V_1$  foi usada a distribuição normal padrão, para  $V_2$  foi usada a distribuição gama com parâmetros 1 e 1, e para  $V_3$  foi usada a distribuição normal com média 50 e desvio padrão 8;

- 4) Uma variável aleatória  $W$ , com distribuição uniforme em  $[0,4]$  e independente de  $V_1$  foi gerada para definir o preditor  $X_1$ , por meio da dicotomização de  $V_1$ , tal que  $X_1 = 1$  se  $W > 1,5$ , e  $X_1 = 0$  em caso contrário;
- 5) O preditor  $X_2$  foi definido como  $X_2 = V_2$ ;
- 6) O preditor  $X_3$  foi definido como o menor inteiro menor ou igual a  $V_3$ ;
- 7) O preditor linear  $XBETA = -1,5 + 1,3x_1 + 0,45x_2 - 0,2\ln(x_3)$  foi utilizado para gerar a probabilidade de sucesso definida por  $p(x) = \frac{e^{XBETA}}{1+e^{XBETA}}$ .
- 8) Foi gerada a variável resposta  $Y$  de acordo com distribuição binomial com parâmetros  $(60, p(x))$ .

A Tabela 1 mostra características das variáveis utilizadas no exemplo deste trabalho. Estatísticas descritivas e correlações são mostradas na Tabela 2 e 3, respectivamente.

Tabela 1 - Descrição das variáveis simuladas.

Variável	Tipo	Natureza	Valores Possíveis
$Y$	Desfecho	Quantitativa	0 ou 60
$X_1$	Preditor	Dicotômico	0 ou 1
$X_2$	Preditor	Quantitativo	(0; 11)
$X_3$	Confundidor	Quantitativo	21 a 81

Tabela 2 - Medidas descritivas.

Variável	Mínimo	Média	Desvio Padrão	Máximo
$Y$	0	17,27580	9,28380	59
$X_1$	0	0,63260	0,48215	1
$X_2$	0,00038	0,97927	0,98136	10,61482
$X_3$	21	50,67380	7,95389	81

Tabela 3 – Matriz de Correlação.

	$Y$	$X_1$	$X_2$	$X_3$
$Y$	1	0,69975	0,59827	-0,39161
$X_1$	0,69975	1	-0,00883	0,01245
$X_2$	0,59827	-0,00883	1	-0,63249
$X_3$	-0,39161	0,01245	-0,63429	1

Interpretando os resultados das Tabelas 1 e 2, pode-se observar que existe grande variabilidade na resposta e nos preditores, e que a resposta está positivamente associada com os preditores  $X_1$  e  $X_2$ , e negativamente associada com  $X_3$ . As Figuras 1, 2 e 3 também ajudam a ilustrar essa variabilidade na resposta  $y$  ao longo dos valores dos preditores  $X_1, X_2$  e  $X_3$ .

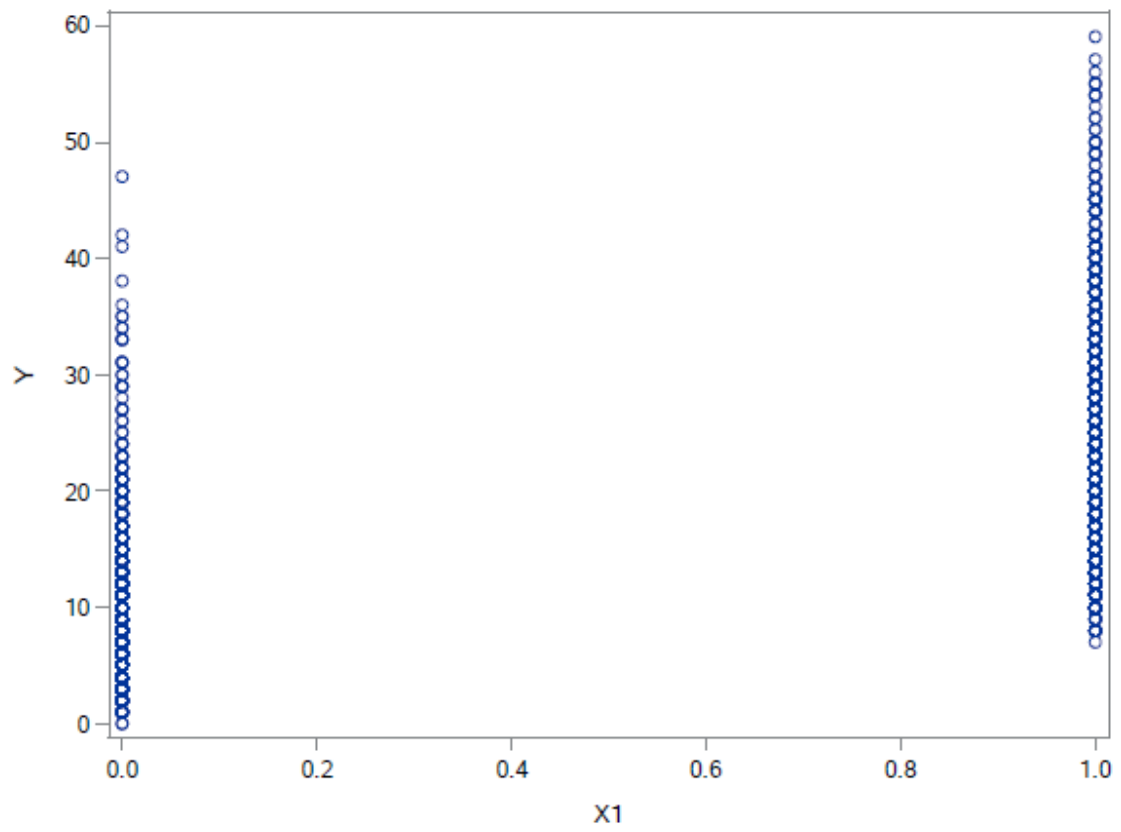


Figura 1: Diagrama de dispersão entre Y e  $X_1$

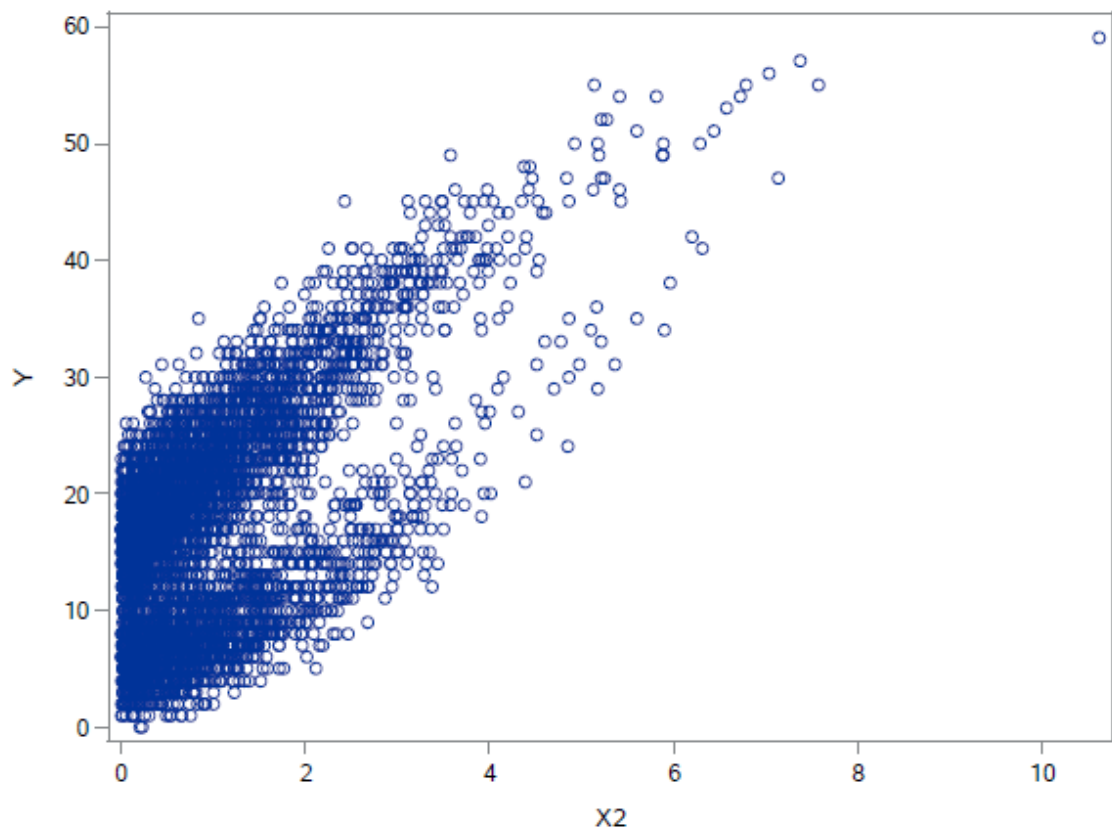


Figura 2: Diagrama de dispersão entre Y e  $X_2$

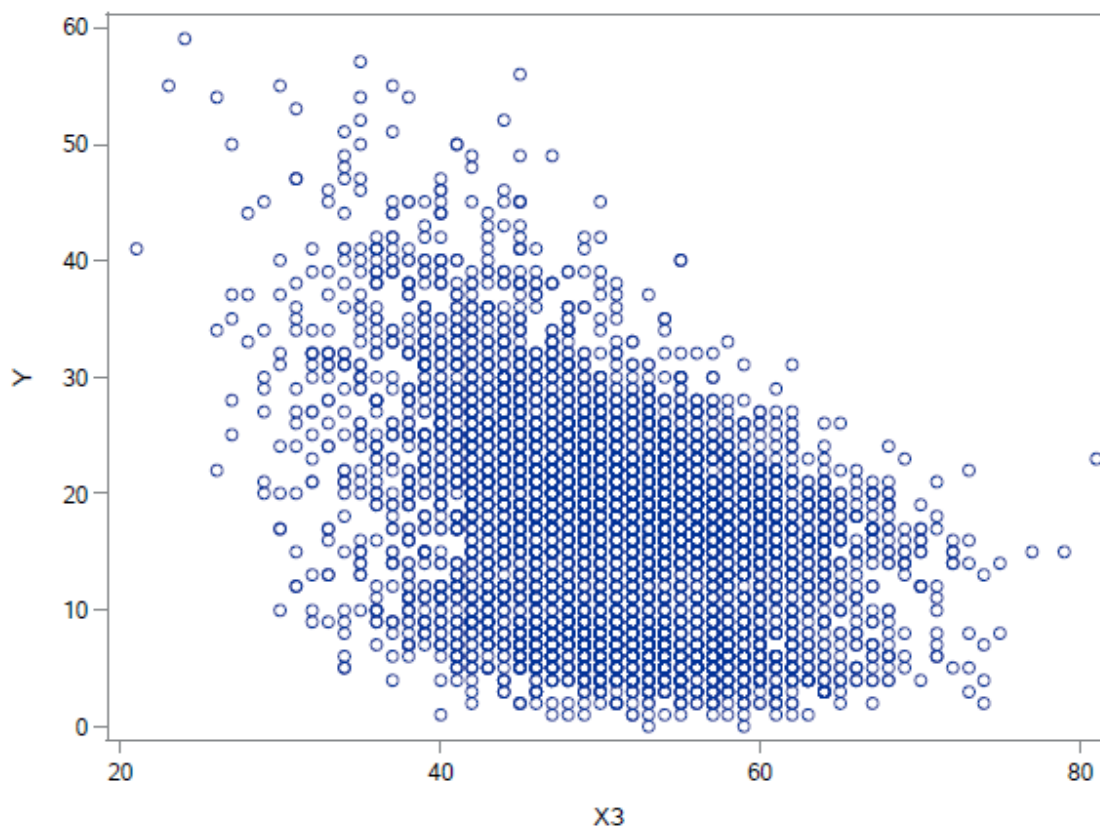


Figura 3: Diagrama de dispersão entre  $Y$  e  $X_3$

## 4.2 Ajuste do modelo

Diferentemente de Bottai, para poder utilizar rotinas computacionais de regressão tradicional, os modelos foram ajustados usando o procedimento PROC QUANTREG do Programa SAS Versão 9.4, por meio da transformação definida em (3). Sendo assim, foram definidos quatro modelos para análise, descritos abaixo. As sintaxes são apresentadas no Anexo.

Modelo 1: Somente  $X_1$

Modelo 2: Somente  $X_2$

Modelo 3:  $X_3 + X_3^2$

Modelo 4:  $X_1 + X_2 + X_3$



Tabela 4 – Estimativas dos parâmetros e de razão de chances para o Modelo 1.

Variável	Quantil	Beta	EP <sub>Beta</sub>	RC <sup>†</sup>	Inferior	Superior
Intercepto	5%	-2,9444	0,0000			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	5%	1,6592	0,0000	5,26	5,26	5,26
Intercepto	25%	-2,1972	0,0000			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	25%	1,2692	0,0000	3,56	3,56	3,56
Intercepto	50%	-1,8718	0,0000			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	50%	1,2528	0,0000	3,50	3,50	3,50
Intercepto	75%	-1,4939	0,0175			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	75%	1,2257	0,0099	3,41	3,34	3,47
Intercepto	95%	-0,8473	0,0582			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	95%	1,3227	0,0789	3,75	3,22	4,38

<sup>†</sup> Estimativa de razão de chances associada com:

$X_1$ : Presente ( $X_1 = 1$ ) versus Ausente ( $X_1 = 2$ )

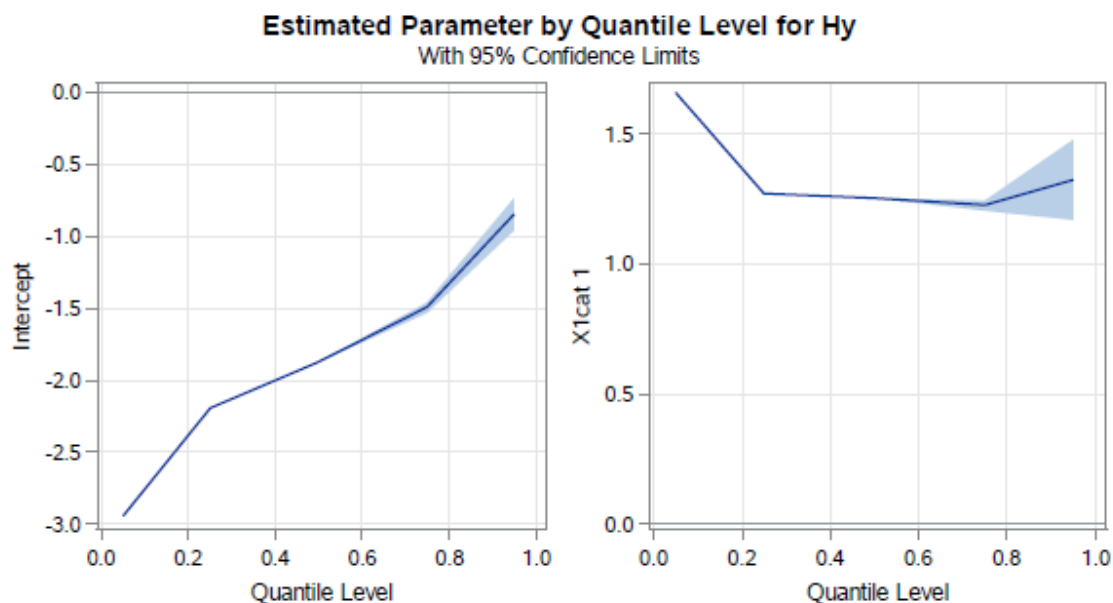


Figura 4: Estimativas e limites de confiança por quantil para as estimativas dos parâmetros do Modelo 1

A Figura 4 mostra as estimativas dos coeficientes do modelo para cada quantil com sua respectiva banda de confiança. Como não inclui o valor zero, evidencia-se que são significativamente diferentes de zero. Para este modelo, a estimativa do coeficiente de regressão é positiva, evidenciando que a

probabilidade do evento ocorrer é maior para  $X_1 = 1$  em relação a  $X_1 = 2$ , e que até o quantil 80 o coeficiente de regressão associado a  $x_1$  diminui, sugerindo que a intensidade da associação entre  $X_1$  e o desfecho  $y$  decresce, aumentando a partir de então. As razões de chances para cada quantil podem ser estimadas de maneira similar à regressão logística para resposta dicotômica.

Assim, para a mediana, por exemplo, a razão de chances de um escore da resposta igual ou superior a mediana para  $X_1 = 1$  versus  $X_1 = 2$  é estimado por  $\exp\{\beta_{0,5}\} = \exp\{1,2528\} = 3,5$ , com intervalo de confiança de 95%, dado por  $\exp(\hat{\beta}_{0,5} \pm 1,96\widehat{EP}_{\hat{\beta}})$ . A interpretação também é a usual, ou seja, estima-se que, a chance média de escore da resposta igual ou superior à mediana é 3,5 vezes maior para indivíduos com  $X_1 = 1$  em relação aos indivíduos com  $X_1 = 2$ , seguindo a mesma lógica de interpretações para os demais percentis estimados visto na Tabela 4.

No Modelo 2, novamente a banda de confiança não contém o valor zero, também evidenciando que o parâmetro de regressão pode ser significativamente diferente ao longo dos quantis da resposta. O coeficiente de regressão diminui até o quantil 25, se mantendo estável nos demais, ou seja, a associação positiva entre  $X_2$  e o desfecho  $y$  tem uma relação decrescente até o quantil 25, mantendo-se estável a partir deste quantil (Tabela 5 e Figura 5).

Tabela 5 – Estimativas dos parâmetros e de razão de chances para o Modelo 2.

Variável	Quantil	Beta	EP <sub>Beta</sub>	RC <sup>†</sup>	Inferior	Superior
Intercepto	5%	-2,9079	0,0343			
$X_2$	5%	0,5695	0,0147	1,77	1,72	1,82
Intercepto	25%	-2,1147	0,0322			
$X_2$	25%	0,4542	0,0176	1,57	1,52	1,63
Intercepto	50%	-1,2277	0,0134			
$X_2$	50%	0,4731	0,0100	1,60	1,57	1,64
Intercepto	75%	0,9391	0,0089			
$X_2$	75%	0,4756	0,0065	1,61	1,59	1,63
Intercepto	95%	-0,6207	0,0106			
$X_2$	95%	0,4732	0,0086	1,61	1,58	1,63

<sup>†</sup> Estimativa de razão de chances associada com:

$X_2$ : Aumento de uma unidade

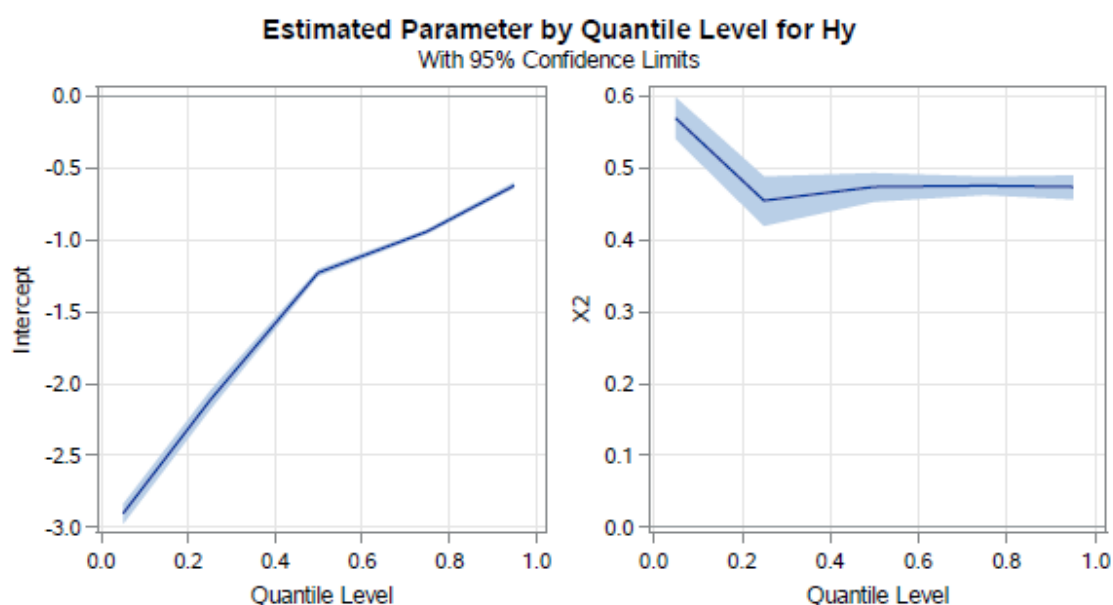


Figura 5: Estimativas e limites de confiança por quantil para as estimativas dos parâmetros do Modelo 2

O Modelo 3 se comporta de maneira diferente dos demais, como pode ser identificado analisando a Tabela 6 e as Figuras 6 e 7, sugerindo que relação entre  $y$  e  $X_3$  pode ser não linear. Ainda, ao contrário dos modelos anteriores, a associação entre o escore da resposta diminui como aumento de  $X_3$ .

Tabela 6 – Estimativas dos parâmetros para o Modelo 3.

Variável	Quantil	Beta	EP <sub>Beta</sub>
Intercepto	5%	2,5440	0,6985
$X_3$	5%	-0,1687	0,0295
$X_3^2$	5%	0,0014	0,0003
Intercepto	25%	3,6134	0,6282
$X_3$	25%	-0,1744	0,0251
$X_3^2$	25%	0,0013	0,0003
Intercepto	50%	2,9839	0,3489
$X_3$	50%	-0,1233	0,0133
$X_3^2$	50%	0,0009	0,0001
Intercepto	75%	4,0531	0,3588
$X_3$	75%	-0,1410	0,0136
$X_3^2$	75%	0,0010	0,0001
Intercepto	95%	6,9847	0,5424
$X_3$	95%	-0,2181	0,0200
$X_3^2$	95%	0,0016	0,0002

† Estimativa de razão de chances associada com:

$X_3$ : Aumento de 5 unidades

$X_3^2$ : Aumento de 25 unidades

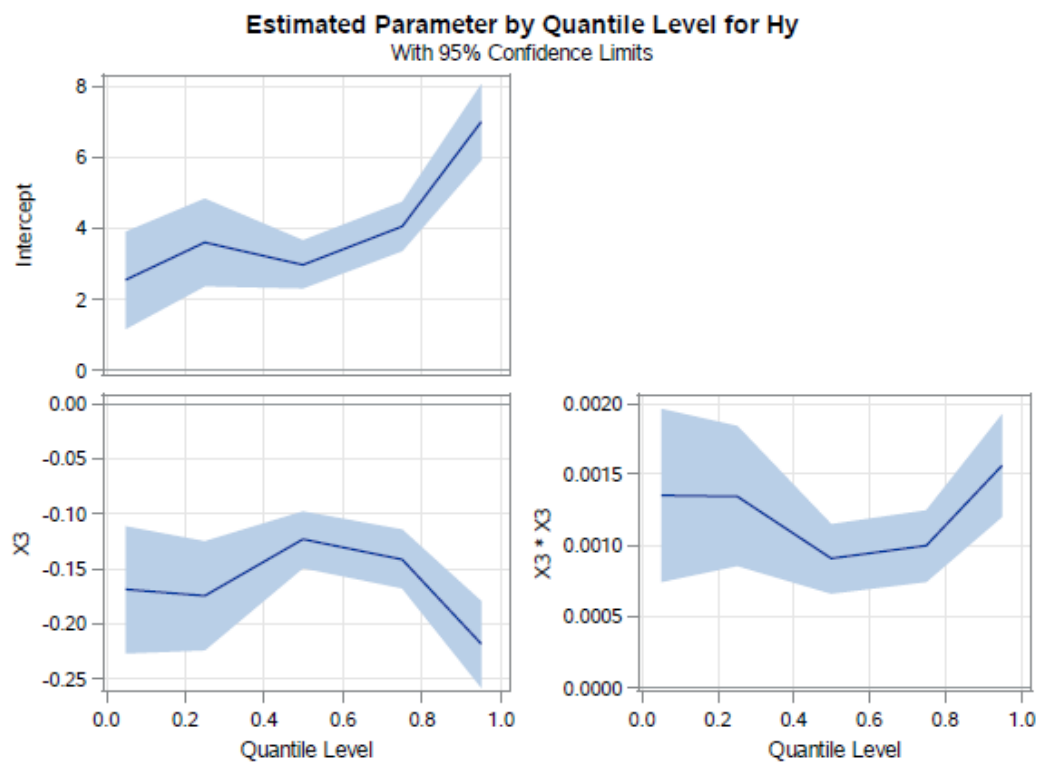


Figura 6: Estimativas e limites de confiança por quantil para as estimativas dos parâmetros do Modelo 3.

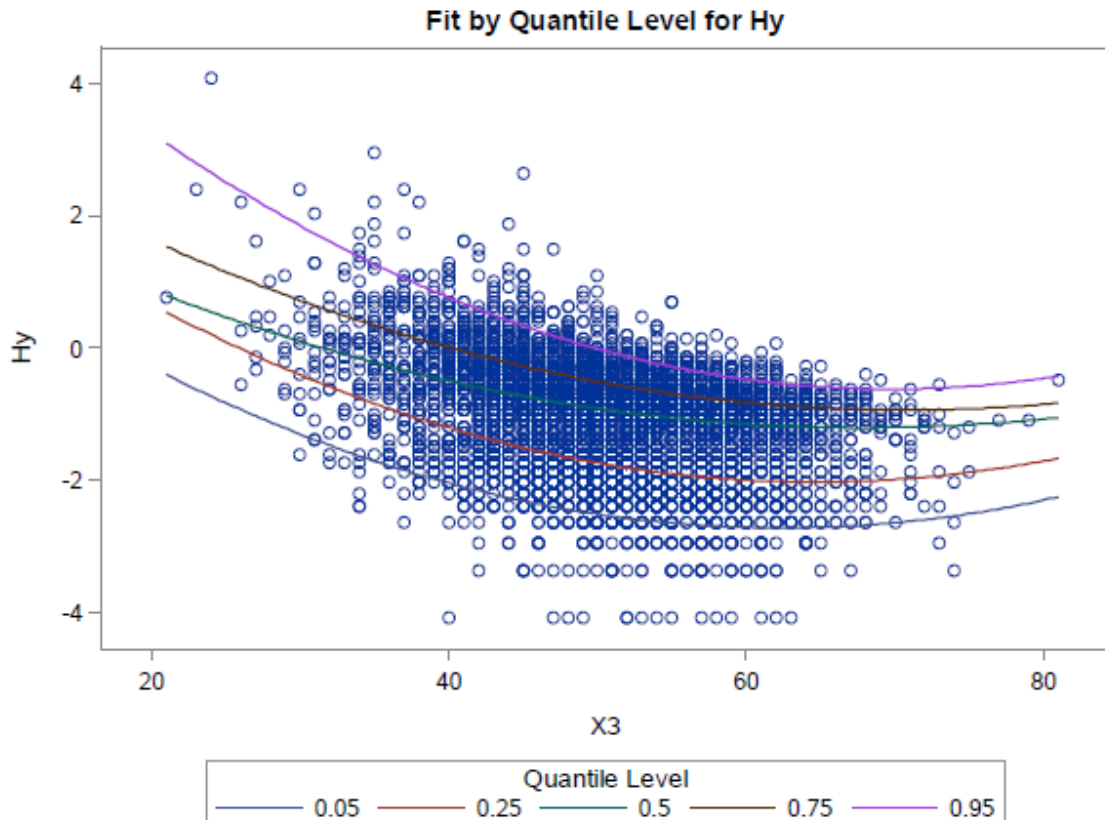


Figura 7: Diagrama de dispersão entre a resposta  $y$  e  $X_3$  e estimativas dos quantis pelo Modelo 3.

A Tabela 7 mostra as estimativas dos parâmetros para o modelo multivariável considerando as variáveis  $X_1$ ,  $X_2$  e  $X_3$ . O termo quadrático  $X_3^2$  foi excluído, pois não foi significativo para nenhum dos quantis ao nível 5%. Por meio deste modelo, estima-se que para cada aumento de 1 unidade em  $X_2$ , aumenta, em média, em 63% a chance de score acima do quantil 5% (IC 95%: 59%-67%). No entanto, analisando as estimativas de razão de chances para os demais percentis, a associação com  $X_2$  diminui à medida que o quantil de resposta aumenta. Para  $X_1$  é possível identificar o mesmo padrão, mas com um decréscimo mais acentuado. Porém, para  $X_3$  a relação é inversa, ou seja, para cada aumento de uma unidade em  $X_3$ , diminui em, em média, em 4% a chance de score da resposta acima do quantil 5% e essa diferença tende a ficar menor, já que é possível identificar que essa associação diminui com o aumento do quantil.

Tabela 7 – Estimativas dos parâmetros e de razão de chances para o Modelo 4.

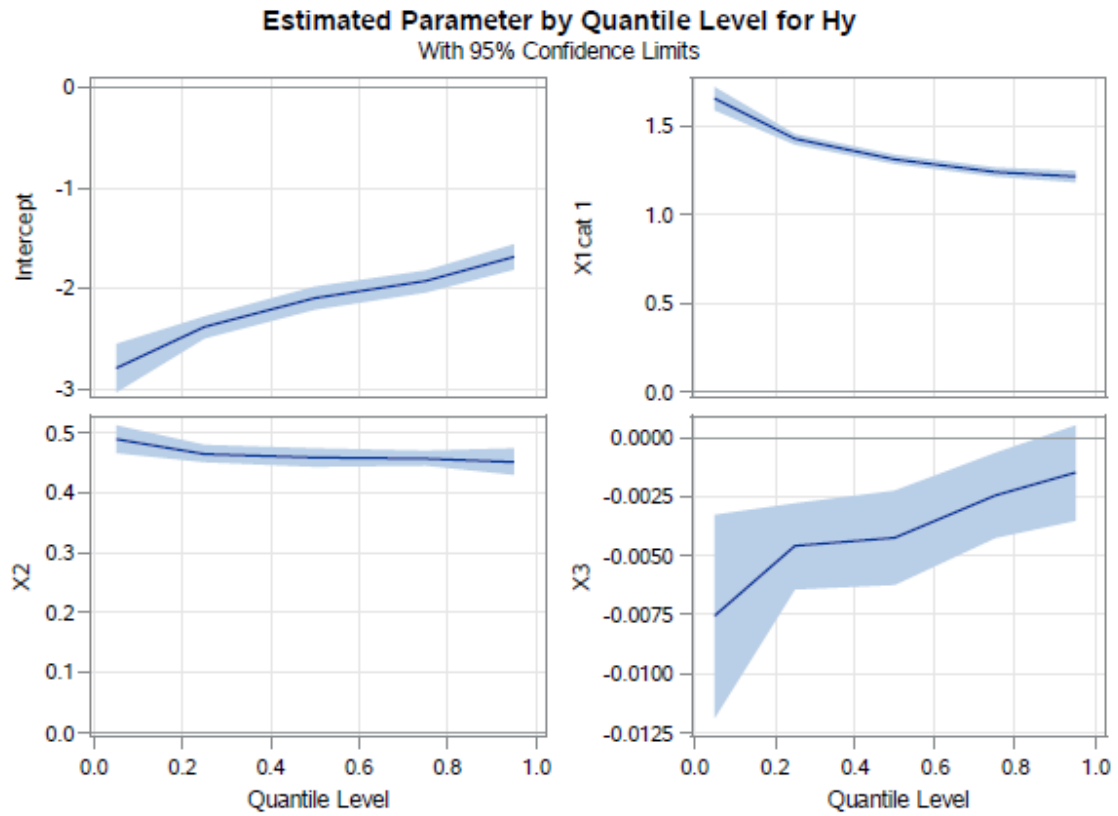
Variável	Quantil	Beta	EP <sub>Beta</sub>	RC <sup>†</sup>	Inferior	Superior
Intercepto	5%	-2,7828	0,1229			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	5%	1,6530	0,0324	5,22	4,90	5,57
$X_2$	5%	0,4891	0,0118	1,63	1,59	1,67
$X_3$	5%	-0,0075	0,0022	0,96	0,94	0,98
Intercepto	25%	-2,3781	0,0556			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	25%	1,4246	0,0153	4,16	4,03	4,28
$X_2$	25%	0,4648	0,0073	1,59	1,57	1,61
$X_3$	25%	-0,0046	0,0009	0,98	0,97	0,99
Intercepto	50%	-2,0944	0,0573			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	50%	1,3101	0,0121	3,71	3,62	3,80
$X_2$	50%	0,4591	0,0077	1,58	1,56	1,61
$X_3$	50%	-0,0042	0,0010	0,98	0,97	0,99
Intercepto	75%	-1,9260	0,0532			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	75%	1,2365	0,0129	3,44	3,36	3,53
$X_2$	75%	0,4572	0,0061	1,58	1,56	1,60
$X_3$	75%	-0,0024	0,0009	0,99	0,98	1,00
Intercepto	95%	-1,6835	0,0650			
$X_1$ ( $X_1 = 1$ vs $X_1 = 2$ )	95%	1,2118	0,0167	3,36	3,25	3,47
$X_2$	95%	0,4518	0,0113	1,57	1,54	1,61
$X_3$	95%	-0,0015	0,0010	0,99	0,98	1,00

<sup>†</sup> Estimativa de razão de chances associada com:

$X_1$ : Presente ( $X_1 = 1$ ) versus Ausente ( $X_1 = 2$ )

$X_2$ : Aumento de uma unidade

$X_3$ : Aumento de 5 unidades



**Figura 8:** Estimativas e limites de confiança por quantil entre  $y$  e  $X_1$ ,  $X_2$  e  $X_3$

Por fim, as Figuras 9, 10, 11, 12 e 13 mostram, respectivamente, os gráficos da distribuição dos resíduos dos quantis 5, 25, 50, 75 e 95 para o Modelo 4. Em todos os casos, nota-se claramente uma distribuição normal.

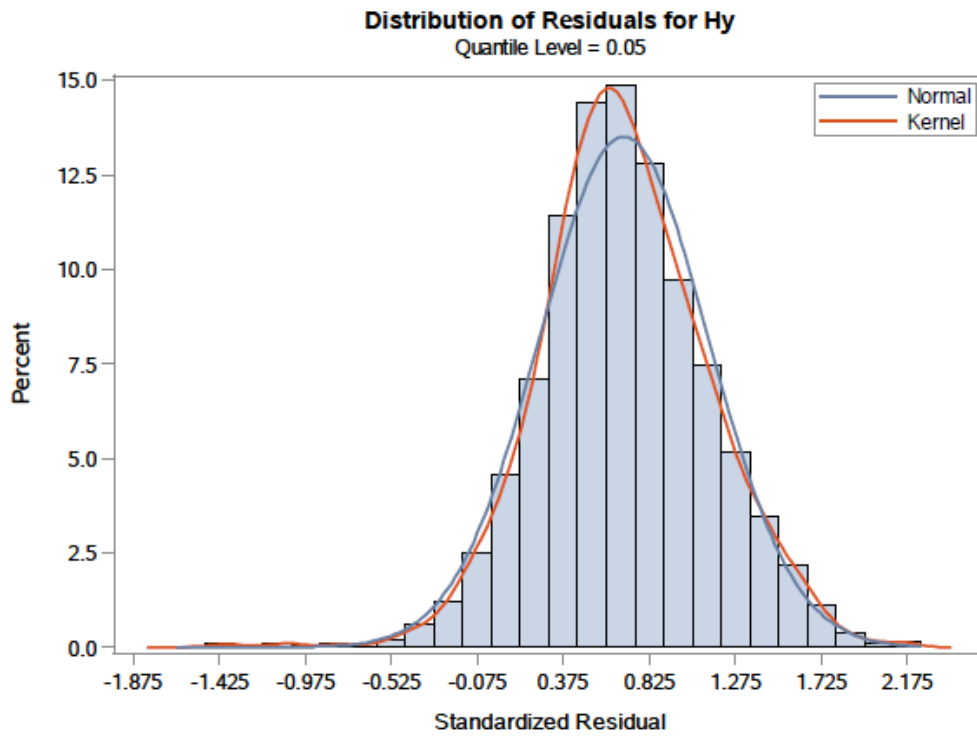


Figura 9: Distribuição dos resíduos do Modelo 4 para o quantil 5.

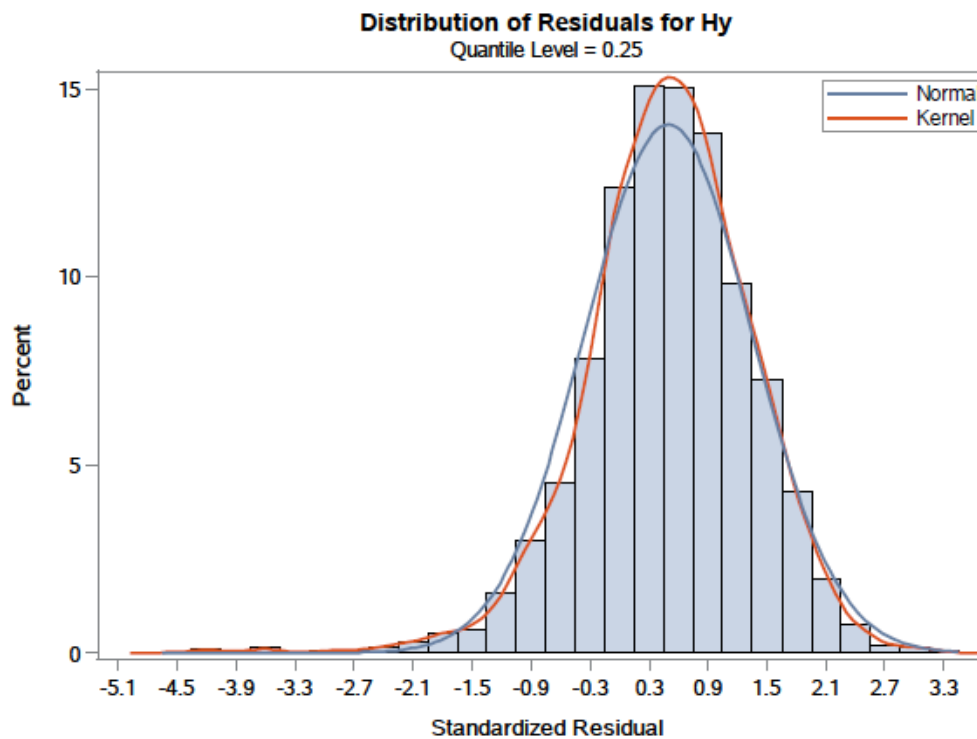


Figura 10: Distribuição dos resíduos do Modelo 4 para o quantil 25



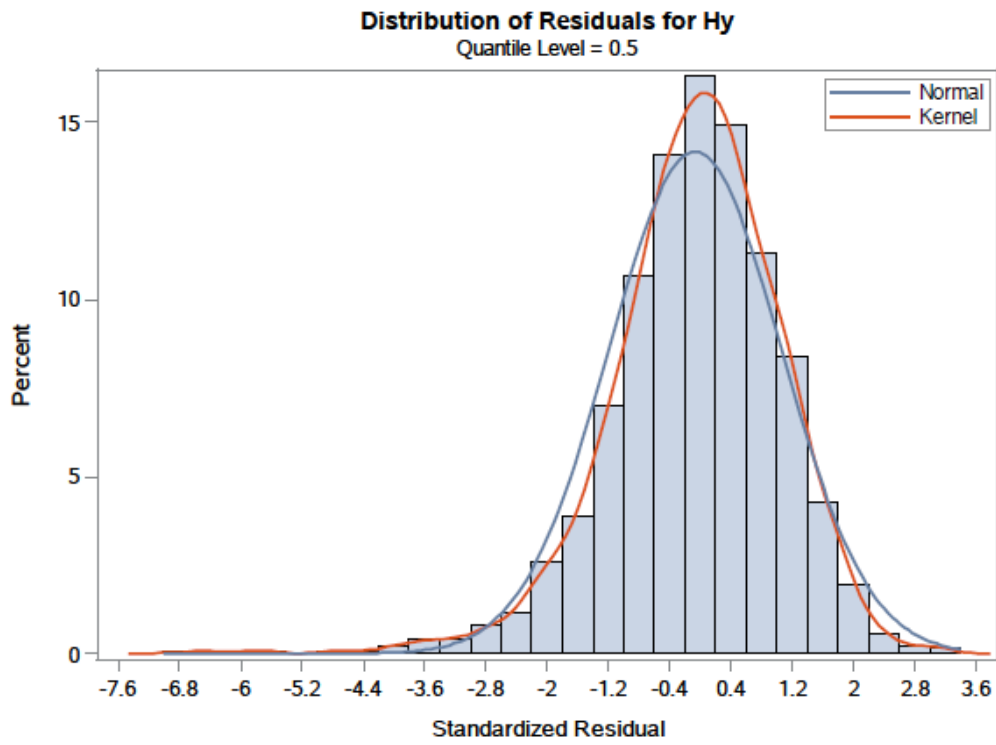


Figura 11: Distribuição dos resíduos do Modelo 4 para o quantil 50.

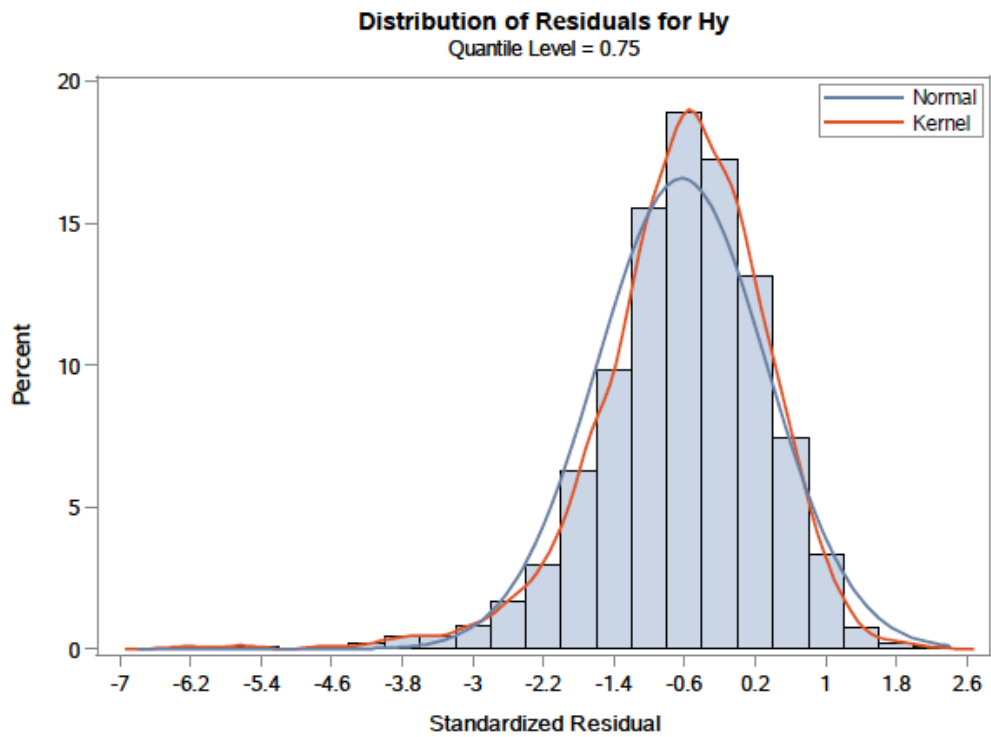


Figura 12: Distribuição dos resíduos do Modelo 4 para o quantil 75.

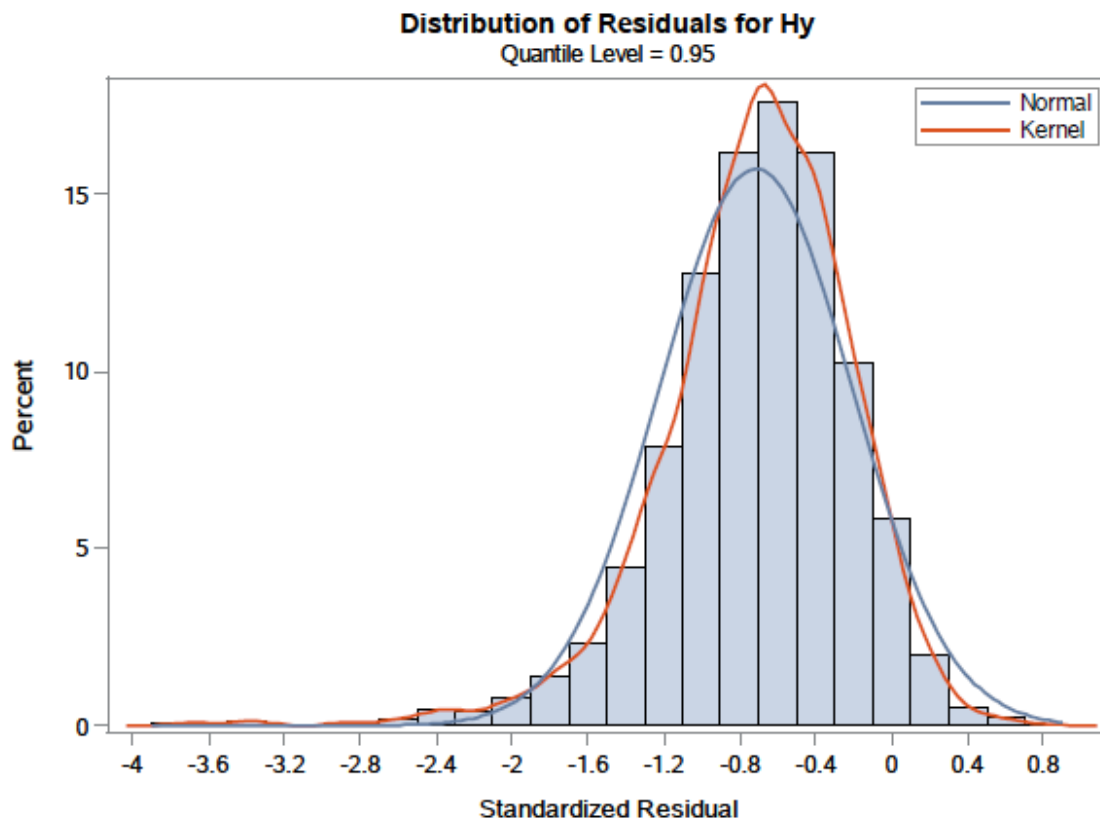


Figura 13: Distribuição dos resíduos do Modelo 4 para o quantil 95.

## 5 CONSIDERAÇÕES FINAIS

Este trabalho mostrou que a abordagem de regressão logística quantílica pode ser uma alternativa importante para análise exploratória de associações no contexto em que a resposta é limitada e tem variabilidade grande. Suas interpretações são realizadas de maneira similar ao modelo de regressão logística tradicional.

Apesar de não existirem procedimentos computacionais específicos, o método de regressão logística quantílica pode ser facilmente realizado por meio de uma transformação na resposta e uso de procedimentos de regressão quantílica tradicionais. Lembrando que podem ser exploradas transformações monótonas, além do logito.

Por fim, este método também pode ser útil para identificar relações não lineares, especialmente quando deseja-se escolher um ponto de corte para resposta.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

Bottai, M., B. Cai, e R. E. McKeown. 2010. Logistic quantile regression for bounded outcomes. *Statistics in Medicine* 29: 309–317.

Koenker, R. 2005. *Quantile Regression*. New York: Cambridge University Press.

## 7 ANEXO – ROTINA SAS PARA AJUSTE DOS MODELOS

```
options ps=58 ls=80 nocenter nodate nonumber formchar='|----|+|---+|=|-\<>*';
libname Tiago 'C:\..\TCC\';
ods pdf file="C:\..\TCC\Tiago\Analises_TCC_Tiago_141115.pdf";

data Modelo;
  set Tiago.Dataset_TCC_Tiago_131115;

  X1cat = 2 - X1;

  Ymin = 0;
  Ymax = 60;

  Hy = log( (Y-Ymin)/(Ymax-Y));
run;
proc means data=Modelo n min mean std max;
  var Y X1 X2 X3;
run;

* MODELOS UNIVARIAVEIS PARA MOSTRAR GRAFICO;
proc quantreg data=Modelo ci=resampling plots=(rdplot ddplot reshistogram);
  title "Modelo Univariavel: X1cat";
  class X1cat;
  model Hy = X1cat / quantile=.05 .25 .5 .75 .95 diagnostics
    leverage(cutoff=8) seed=1268 plot=quantplot;
  id ID;
run;

proc quantreg data=Modelo ci=resampling plots=(rdplot ddplot reshistogram);
  title "Modelo Univariavel: X2";
  model Hy = X2 / quantile=.05 .25 .5 .75 .95 diagnostics
    leverage(cutoff=8) seed=1268 plot=quantplot;
  id ID;
run;

proc quantreg data=Modelo ci=resampling plots=(rdplot ddplot reshistogram);
  title "Modelo Univariavel: X3";
  model Hy = X3 / quantile=.05 .25 .5 .75 .95 diagnostics
    leverage(cutoff=8) seed=1268 plot=quantplot;
  id ID;
run;

proc quantreg data=Modelo ci=resampling plots=(rdplot ddplot reshistogram);
  title "Modelo Univariavel: X3 + X3*X3";
  model Hy = X3 X3*X3 / quantile=.05 .25 .5 .75 .95 diagnostics
    leverage(cutoff=8) seed=1268 plot=quantplot;
  id ID;
run;

title "Modelo Multivariavel: X1cat + X2 + X3 + X3*X3";
proc quantreg data=Modelo ci=resampling plots=(rdplot ddplot reshistogram);
  class X1cat ;
  model Hy = X1cat X2 X3 X3*X3 / quantile=.05 .25 .5 .75 .95 diagnostics
    leverage(cutoff=8) seed=1268 plot=quantplot;
  id ID;
run;
title;

title "Modelo Multivariavel: X1cat + X2 + X3";
ods output ParameterEstimates=ParamEstimates;
proc quantreg data=Modelo ci=resampling plots=(rdplot ddplot reshistogram);
  class X1cat ;
  model Hy = X1cat X2 X3 / quantile=.05 .25 .5 .75 .95 diagnostics
    leverage(cutoff=8) seed=1268 plot=quantplot;
  id ID;
run;
```