



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
CURSO DE GRADUAÇÃO EM

LUCAS PEREIRA DO AMARAL
PEDRO HENRIQUE RODRIGUES DE LIMA

MODELOS DE REGRESSÃO BETA

FORTALEZA

2022

RESUMO

Esse relatório consiste no estudo do modelo de regressão beta, o qual é uma alternativa para a modelagem de variáveis resposta restritas a um intervalo unitário (0,1), como por exemplo, taxas e proporções. O relatório foi baseado no artigo de Ferrari e Cribari-Neto (2004), no qual foi proposto esse modelo em questão, e na dissertação de mestrado de Oliveira (2004). Essencialmente, o modelo de regressão beta é utilizada para modelar uma variável resposta com distribuição beta, sendo utilizada uma parametrização indexada pela média e por um parâmetro de dispersão. Ademais, foi desenvolvida uma análise num contexto prático, a fim de se averiguar como é o comportamento do modelo.

Palavras-chave: modelos de regressão; distribuição beta; modelagem; taxas ; proporções.

LISTA DE FIGURAS

Figura 1 – Gráficos da densidade de beta para diferentes combinações dos valores de p e q.	7
Figura 2 – Densidades beta para diferentes combinações de μ e ϕ (Ferrari, S. L. P. e Cribari-Neto, F.(2004)).	11
Figura 3 – Gráfico da perda de conversão x índice das observações.	22
Figura 4 – Gráficos para o diagnóstico do modelo.	23
Figura 5 – Envelope simulado meio-normal para os resíduos componentes do desvio.	24

LISTA DE TABELAS

Tabela 1 – Dados do estudo de oxidação da Amônia	21
Tabela 2 – Estimativas dos parâmetros	22
Tabela 3 – Estimativas dos parâmetros sem a variável x_3	23

SUMÁRIO

1	INTRODUÇÃO	5
2	DISTRIBUIÇÃO BETA	6
2.1	Estimação dos parâmetros	8
2.1.1	<i>Método de Máxima Verossimilhança</i>	8
3	MODELO DE REGRESSÃO BETA	10
3.1	O Modelo	11
3.2	Funções de Ligação	12
3.3	Função Escore e Matriz de Informação de Fisher	13
3.4	Estimação dos Parâmetros	16
3.4.1	<i>Método de Newton-Raphson</i>	17
3.5	Inferências de Segunda Ordem	17
3.5.1	<i>Teste de Wald</i>	17
3.6	Técnicas de Diagnóstico	18
3.6.1	<i>Resíduo Componente do Desvio</i>	18
3.6.2	<i>Gráficos de Envelope Simulado Meio-Normal</i>	19
3.6.3	<i>Alavancagem Generalizada</i>	19
3.6.4	<i>Influência</i>	19
4	APLICAÇÃO	21
5	CONCLUSÃO	25
	REFERÊNCIAS	26
	APÊNDICE A –INFORMAÇÃO DE FISHER	27

1 INTRODUÇÃO

O modelo de regressão beta foi proposto por Ferrari e Cribari-Neto [2004. Beta regression for modeling rates and proportions. J. Appl. Statist. 31, 799–815] objetivando modelar variáveis contínuas no intervalo $(0,1)$, por exemplo, um caso em que a variável resposta é dada por taxas e proporções.

Ademais, como proposto pelos autores citados acima, a ideia da modelagem consiste em utilizar a resposta assumindo distribuição beta juntamente com uma parametrização da lei beta reajustada pela média e por um parâmetro de dispersão. Entretanto, ao fazer a estimação dos parâmetros via função de verossimilhança, depara-se com o fato de que não há solução analítica, com isso, faz-se o uso de métodos numéricos para a obtenção dos estimadores.

À luz dessas considerações, coloca-se, em pauta, as questões envolvendo a modelagem utilizando a distribuição beta, a qual será abordada no decorrer do trabalho.

2 DISTRIBUIÇÃO BETA

A família de distribuições beta é composta de todas as distribuições com função densidade de probabilidade (f.d.p.) da forma:

$$f(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1, \quad p > 0, \quad q > 0, \quad (2.1)$$

Com média e variância dadas por:

$$E[Y] = \frac{p}{p+q} \quad (2.2)$$

e

$$Var[Y] = \frac{pq}{(p+q)^2(p+q+1)} \quad (2.3)$$

Pham-Gia (1994) mostrou alguns limites para a variância de y . Essencialmente, ele demonstrou que $Var[Y] < 1/4$, e se a densidade for unimodal, o que acontece quando $p > 1$ e $q > 1$, a $Var[Y] < 1/12$. Caso a densidade seja em formato de U, ou seja, $p < 1$ e $q < 1$, ele provou que $1/12 < Var[Y] < 1/4$.

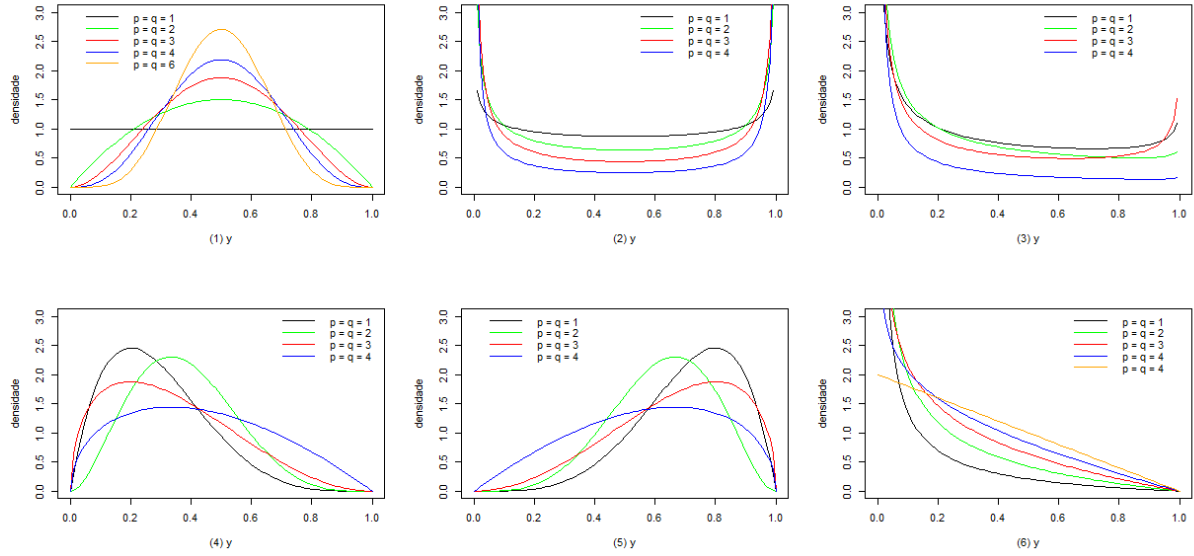
Ademais, $\Gamma(p)$ é a função gama avaliada num dado ponto p , assim,

$$\Gamma(p) = \int_0^\infty y^{p-1} e^{-y} dy.$$

A função densidade de probabilidade dada em (2.1) é a forma padrão da distribuição em questão. Também é possível notar que a distribuição uniforme é um caso particular quando $q = p = 1$.

Além disso, o modelo beta não possui a estrutura dos modelos de locação-escala, pois ambos os parâmetros p e q são parâmetros de forma, conferindo bastante flexibilidade quanto ao formato da densidade. Pode - se observar isso na figura 1, para várias combinações de p e q .

Figura 1 – Gráficos da densidade de beta para diferentes combinações dos valores de p e q .



Os gráficos 1 (1) e 1 (2) apresentam as densidades simétricas ($p = q$), em (1) são unimodais e em (2) são em formato de U ($p, q < 1$). Os gráficos 1 (3), (4) e (5) denotam densidades assimétricas, em (3) tem-se novamente o formato de U, já em (4) e (5) ocorre a representação unimodal e assimétrica à direita e à esquerda, respectivamente. Por fim, no gráfico 1 (6) ocorre a densidade em formato J, onde $(p - 1)(q - 1) < 0$.

De forma geral, essas características podem ser observadas de forma analítica por meio do coeficiente de assimetria e pelo coeficiente de curtose. Sendo o coeficiente de assimetria dado por

$$\lambda_1 = \frac{2(q-p)}{p+q+2} \sqrt{\frac{p+q+1}{pq}}.$$

Sabe-se que esta medida caracteriza como e quanto a distribuição se afasta da condição de simetria. Com isso, se $p = q$, então $\lambda_1 = 0$, tem-se simetria. Se $q > p$, $\lambda_1 > 0$, nota-se assimetria à direita. De forma análoga, caso $q < p$, $\lambda_1 < 0$, conferindo, assim, assimetria à esquerda.

Ademais, tem-se o coeficiente de curtose, o qual é utilizado para caracterizar o formato da distribuição em relação ao seu achatamento. Sendo dado por

$$\lambda_2 = \frac{3(p+q+1)[2(p+q)^2 + pq(p+q-6)]}{pq(p+q+2)(p+q+3)}.$$

Por meio da medida acima, nota-se $\lambda_2 < 3$ para distribuições platycúrticas, $\lambda_2 = 3$ para mesocúrticas e $\lambda_2 > 3$ para leptocúrticas. Em especial, no caso simétrico, em que $p = q = \lambda$, $\lambda_2 \xrightarrow{3}$ quando λ cresce, fazendo a f.d.p. da beta se aproximar da distribuição normal com média $1/2$ e variância $1/[4(2\lambda + 1)]$.

2.1 Estimação dos parâmetros

Nesta seção, serão obtidas as estimativas para os parâmetros da distribuição beta, por meio de dois métodos bastante tradicionais. Em primeira análise, será abordada a estimação via método de máxima verossimilhança, no qual os estimadores são obtidos através da maximização da função de verossimilhança. Em segunda análise, coloca-se, em pauta, o método dos momentos, o qual consiste em igualar os momentos amostrais aos correspondentes momentos populacionais.

2.1.1 Método de Máxima Verossimilhança

Sejam Y_1, \dots, Y_n uma amostra aleatória de tamanho n da variável aleatória $Y_i \sim \text{Beta}(\theta)$, sendo $\theta = (p, q) \in \Theta = \mathbb{R}^+ \times \mathbb{R}^+$, sendo Θ o espaço paramétrico. Logo, a função de verossimilhança de θ é dada por

$$L(\theta) = L(p, q) = \prod_{t=1}^n \frac{y_t(1-y_t)^{q-1}}{\mathbb{B}(p, q)} = [\mathbb{B}(p, q)]^{-n} \prod_{t=1}^n y_t^{p-1} \prod_{t=1}^n (1-y_t)^{q-1}, \quad (2.4)$$

Na qual $\mathbb{B}(p, q)$ é a função beta dada por

$$\mathbb{B}(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}. \quad (2.5)$$

Com isso, utilizando o logarítmo natural em ambos os lados da equação (2.4), tem-se

$$l(p, q) = \log L(p, q) = -n \log(\mathbb{B}(p, q)) + (p-1) \sum_{t=1}^n \log y_t + (q-1) \sum_{t=1}^n \log(1-y_t)$$

Derivando $l(p, q)$ em relação a cada parâmetro e igualando à zero, tem-se as equações de máxima verossimilhança. Assim,

$$\psi(\hat{p}) - \psi(\hat{p} + \hat{q}) = \frac{1}{n} \log \sum_{t=1}^n \log y_t \quad \text{e} \quad \psi(\hat{q}) - \psi(\hat{p} + \hat{q}) = \frac{1}{n} \log \sum_{t=1}^n \log(1-y_t), \quad (2.6)$$

sendo $\psi(\hat{\lambda})$ é a função digama dada por

$$\psi(\hat{\lambda}) = \frac{\log \Gamma(\hat{\lambda})}{\hat{\lambda}} = \frac{\Gamma'(\hat{\lambda})}{\Gamma(\hat{\lambda})}, \quad \lambda > 0. \quad (2.7)$$

Nota-se que não existem soluções analíticas explícitas para a estimação dos parâmetros da distribuição beta pelas equações de máxima verossimilhança. Neste caso, pode-se obter os estimadores por métodos numéricos, por exemplo, pelos métodos de Newton-Raphson ou Escore de Fisher.

Sob condições gerais de regularidade (ver, Sen e Singer, 1993, Capítulo 5), quando a amostra é suficientemente grande, tem-se

$$\begin{pmatrix} \hat{p} \\ \hat{q} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} p \\ q \end{pmatrix}, K^{-1}(p, q) \right), \quad (2.8)$$

de forma aproximada, na qual \hat{p} e \hat{q} são os estimadores de máxima verossimilhança. Já a matriz de covariâncias assintótica dos estimadores $K^{-1}(p, q)$ (ver Apêndice A) é

$$K^{-1}(p, q) = n^{-1} [\psi'(p)\psi'(q) - \psi'(p+q)\psi'(p) + \psi'(q)]^{-1} \begin{pmatrix} \psi'(q) - \psi'(p+q) & \psi'(p+q) \\ \psi'(p+q) & \psi'(p) - \psi'(p+q) \end{pmatrix},$$

sendo $\psi'(\lambda)$ a função trigama, dada por

$$\psi'(\lambda) = \frac{d\psi(\lambda)}{d\lambda} = \frac{d^2\Gamma(\lambda)}{d\lambda^2}. \quad (2.9)$$

Assim, por meio das equações (2.7) e (2.8), mostra-se que os estimadores de máxima verossimilhança são assintoticamente não viciados e consistentes.

3 MODELO DE REGRESSÃO BETA

A densidade de beta é dada pela equação (2.1), contudo, modelos de regressão usualmente utilizam uma estrutura para modelar a média da variável resposta e um parâmetro de precisão. Com isso, Ferrari e Cribari-Neto (2004) propuseram a reparametrização da densidade de beta.

Assim, seja $\mu = p/(p + q)$ e $\phi = p + q$, isto é, $p = \mu\phi$ e $q = (1 - \mu)\phi$. Logo, a média e a variância podem ser reescritas como:

$$E[Y] = \mu$$

e

$$Var[Y] = \frac{V(\mu)}{1 + \phi}$$

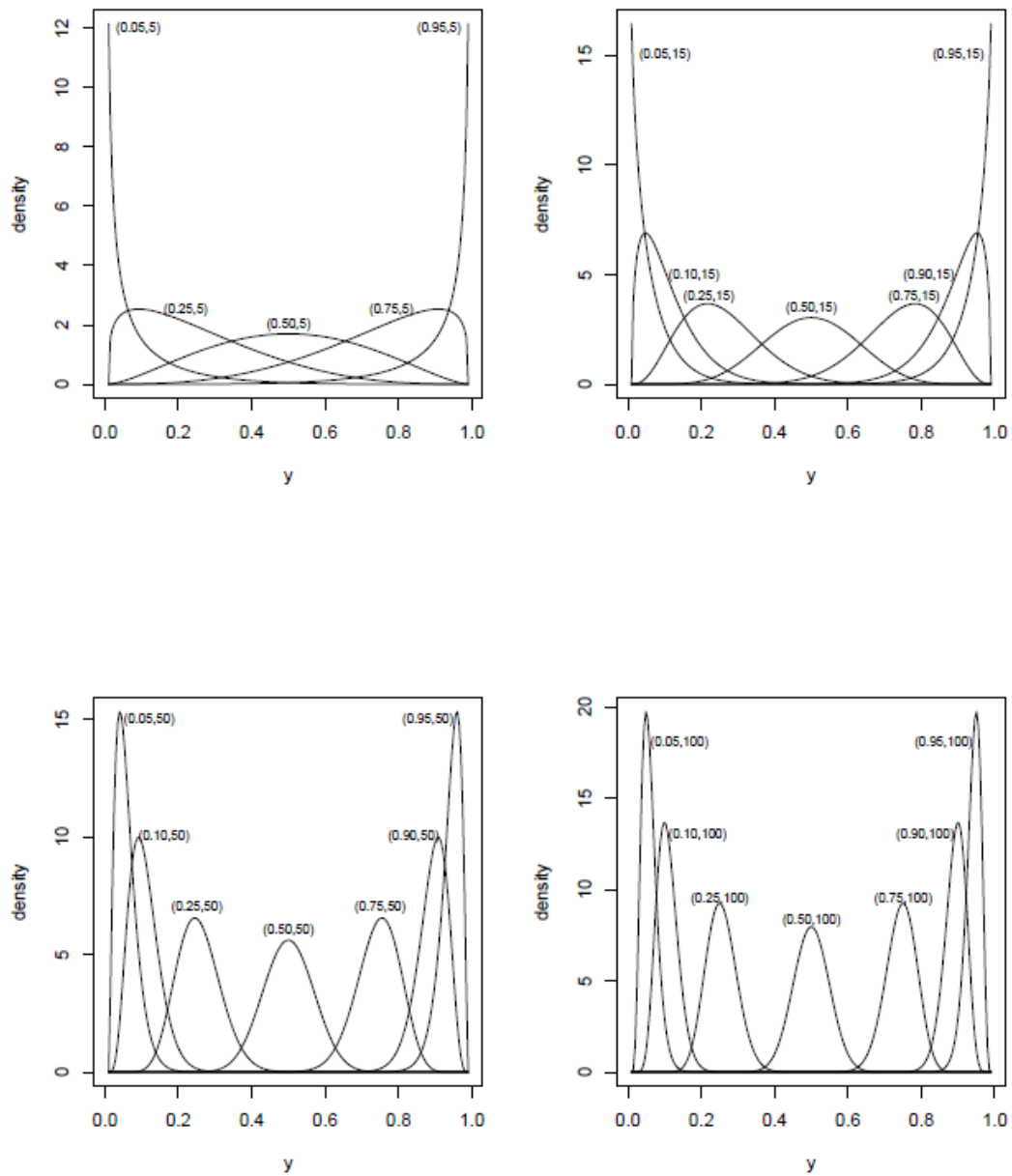
Onde $V(\mu) = \mu(1 - \mu)$, com isso, tem-se que μ sendo a média da variável resposta e ϕ é colocado como um parâmetro de precisão, na ideia de que, para μ fixado, quanto maior for o valor de ϕ , menor será a variância de y . Portanto, a densidade da beta pode ser reescrita como:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi-1} (1 - y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad (3.1)$$

em que $0 < \mu < 1$ e $\phi > 0$.

A fim de verificar o comportamento da densidade com outra parametrização, na Figura 2 são apresentadas algumas densidades com diferentes valores de μ e ϕ . Nota-se que a forma da densidade varia de acordo com os valores dados para os parâmetros. Quando $\mu = 1/2$, pode-se haver simetria, e assimetria quando $\mu \neq 1/2$. Ademais, fixando-se μ , a dispersão da distribuição irá diminuir quando ϕ aumentar. Além disso, tem-se também, densidades em formato de J e J invertido, uniformes e em formato de U, simétrica e assimétricas.

Figura 2 – Densidades beta para diferentes combinações de μ e ϕ (Ferrari, S. L. P. e Cribari-Neto, F.(2004)).



3.1 O Modelo

Apesar de o modelo ser utilizado para uma variável resposta que assume valores no intervalo $(0,1)$, ele pode ser utilizado em situações mais gerais. Essencialmente, se a resposta assumir valores em (a,b) , sendo a e b ($a < b$), duas constantes conhecidas, modela-se, nesse caso, $(y-a)/(b-a)$, ao invés de modelar y diretamente.

Sejam y_1, y_2, \dots, y_n variáveis aleatórias independentes, em que cada y_t , $t = 1, \dots, n$ tem densidade como em (3.1), tendo média μ_t e parâmetro de precisão ϕ . Assim, o modelo de regressão beta é por (3.1), juntamente com o componente sistemático

$$g(\mu_t) = \sum_{i=1}^k x_{ti} \beta_i = \eta_t, \quad (3.2)$$

Onde $\eta_t = x_t^T \beta$ é o preditor linear, $\beta = (\beta_1, \dots, \beta_k)^T$ é o vetor de parâmetros a serem estimados ($\beta \in \mathbb{R}^k$), $x_t^T = (x_{t1}, \dots, x_{tk})$ representa os valores de k ($k < n$) variáveis explicativas, sendo elas fixas e conhecidas. E $g(\cdot)$ é uma função estritamente monótna e duplamente diferenciável que transforma valores do intervalo $(0,1)$ em \mathbb{R} , denominada de função de ligação. É interessante notar que a variância de y_t é uma função de μ_t e, pelo modelo (3.2), das variáveis explicativas. Portanto, variáveis resposta y_t de variâncias não constantes são naturalmente acomodadas no modelo, não há a necessidade de, se ter ou supor, homocedasticidade.

3.2 Funções de Ligação

A função de ligação que será utilizada no modelo tem papel primordial não somente no ajuste, mas também na interpretação dos parâmetros, ou seja, é imprescindível para que o modelo seja, de fato, eficaz, não somente no contexto teórico, mas numa conjuntura prática.

A função de ligação $g(\cdot)$ pode ser a especificação logito, a função probito que é a função de distribuição acumulada de uma variável aleatória normal padrão, pode ser também a ligação complemento log-log, a ligação log-log, a ligação cauchit, entre outras. Por uma questão de usualidade, pois será utilizada numa aplicação posteriormente, analisando a logito

$$g(\mu_t) = \log\left(\frac{\mu_t}{1 - \mu_t}\right) = x_t^T \beta, \quad t = 1, \dots, n$$

Em que $\log\left(\frac{\mu_t}{1 - \mu_t}\right)$ é a especificação logito.

Assim, o modelo pode ser escrito como

$$\frac{\mu_t}{1 - \mu_t} = \exp(x_t^T \beta)$$

ou

$$\mu_t = \frac{\exp(x_t^T \beta)}{1 + \exp(x_t^T \beta)}. \quad (3.3)$$

A equação (3.3) é a função inversa de $g(\mu_t)$. Para a interpretação dos parâmetros do modelo, supõe-se que o valor da i -ésima variável regressora é aumentado por c unidades e

todas as outras variáveis explicativas fixadas. Sendo μ^\dagger a média de y sob esse novo valor das covariadas, uma vez que μ denota a média de y sob o valor original das covariadas. Tem-se, então,

$$\frac{\mu^\dagger}{1 - \mu^\dagger} = \exp(x_{t1}\beta_1 + \dots + (x_{ti} + c)\beta_i + \dots + x_{tk}\beta_k).$$

Assim, verifica-se

$$\exp(c\beta_i) = \frac{\mu^\dagger/(1 - \mu^\dagger)}{\mu/(1 - \mu)}.$$

Ou seja, $\exp(c\beta_i)$ é a razão de chances (*odds ratio*).

3.3 Função Escore e Matriz de Informação de Fisher

O logaritmo natural da função de verossimilhança baseada em uma amostra de n observações independentes é dado por

$$l(\beta, \phi) = \sum_{t=1}^n l_t(\mu_t, \phi), \quad (3.4)$$

em que

$$l_t(\mu_t, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t)\phi) + (\mu_t \phi - 1) \log y_t + \{(1 - \mu_t)\phi - 1\} \log(1 - y_t)$$

sendo $\mu_t = g^{-1}(\eta_t)$ definido em (3.2). Para que a função escore seja obtida, deriva-se o logaritmo da função de verossimilhança em relação aos parâmetros. Assim, para $i = 1, \dots, k$,

$$\frac{\partial l(\beta, \phi)}{\partial \beta_i} = \sum_{t=1}^n l_t \frac{\partial l_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_i} \quad (3.5)$$

com

$$\frac{\partial l_t(\mu_t, \phi)}{\partial \mu_t} = \phi \left[\log \frac{y_t}{1 - y_t} - \{\psi(\mu_t \phi) - \psi((1 - \mu_t)\phi)\} \right], \quad (3.6)$$

onde $\psi(\cdot)$ é a função digama, definida em (2.7). Ademais, $d\mu_t/d\eta_t = 1/g'(\mu_t)$ e $\partial \eta_t/\partial \beta_i = x_{ti}$.

Além disso, fazendo-se $y_t^* = \log\{y_t/(1 - y_t)\}$ e $\mu_t^* = \psi(\mu_t \phi) - \psi((1 - \mu_t)\phi)$.

Logo, pode-se reescrever (3.6) como

$$\frac{\partial l_t(\mu_t, \phi)}{\partial \mu_t} = \phi(y_t^* - \mu_t^*),$$

finalmente, (3.5) é dada por

$$\frac{\partial l(\beta, \phi)}{\partial \beta_i} = \phi \sum_{t=1}^n (y_t^* - \mu_t^*) \frac{1}{g'(\mu_t)} x_{ti}. \quad (3.7)$$

É notório que supondo condições usuais de regularidade (ver, Sen e Singer, 1993, Capítulo 7), o valor esperado de $\partial l_t(\mu_t, \phi)/\mu_t$ é igual a zero. Contudo,

$$E \left[\frac{\partial l_t(\mu_t, \phi)}{\mu_t} \right] = 0 \Leftrightarrow E[\phi(y_t^* - \mu_t^*)] = 0 \Leftrightarrow E[y_t^*] = \mu_t^*$$

À luz dessas considerações, nota-se que a esperança da variável transformada y_t^* acaba se igualando a μ_t^* .

Com isso, a função escore para β , $U_\beta(\beta, \phi)$, um vetor coluna de dimensão k , pode ser denotada pela forma matricial

$$U_\beta(\beta, \phi) = \phi X^T T(y^* - \mu^*), \quad (3.8)$$

onde X é uma matriz ($n \times k$), na qual a t -ésima linha é x_t^T , $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$, $y^* = (y_1^*, \dots, y_n^*)^T$ e $\mu^* = (\mu_1^*, \dots, \mu_n^*)^T$.

Para o parâmetro de precisão ϕ , o processo é análogo, assim, a derivada de 1ª ordem do logaritmo da função de verossimilhança em relação a ϕ é

$$\frac{\partial(\beta, \phi)}{\partial \phi} = \sum_{t=1}^n \frac{l_t(\mu_t, \phi)}{\partial \phi}, \quad (3.9)$$

com

$$\frac{\partial l_t(\mu_t, \phi)}{\partial \phi} = \mu_t \left[\log \frac{y_t}{1-y_t} - \psi(\mu_t \phi) + \psi((1-\mu_t)\phi) \right] + \log(1-y_t) - \psi((1-\mu_t)\phi) + \psi(\phi) \quad (3.10)$$

Portanto, a função escore para ϕ é

$$U_\phi(\beta, \phi) = \sum_{t=1}^n \{ \mu_t(y_t^* - \mu_t^*) + \log(1-y_t) - \psi((1-\mu_t)\phi) + \psi(\phi) \}, \quad (3.11)$$

sendo $U_\phi(\beta, \phi)$ um escalar.

Para obter a matriz de informação de Fisher, por meio de (3.5), as derivadas de 2ª ordem de $l(\beta, \phi)$ em relação a β_i e β_j são dadas por

$$\begin{aligned} \frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \beta_j} &= \sum_{t=1}^n \frac{\partial}{\partial \mu_t} \left(\frac{\partial l_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_j} x_{ti} = \\ &= \sum_{t=1}^n \left\{ \frac{\partial^2 l_t(\mu_t, \phi)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} + \frac{\partial l_t(\mu_t, \phi)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_t} \right) \right\} \frac{d\mu_t}{d\eta_t} x_{ti} x_{tj} \end{aligned} \quad (3.12)$$

Mas como provado anteriormente, $E[\partial l_t(\mu_t, \phi)/\partial \mu_t] = 0$, com isso, tem-se

$$\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \beta_j} = \sum_{t=1}^n E \left(\frac{\partial^2 l_t(\mu, \phi)}{\partial \mu_t^2} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2 x_{ti} x_{tj}$$

De (3.6), tem-se

$$\frac{\partial^2 l_t(\mu, \phi)}{\partial \mu_t^2} = -\phi^2 \{ \psi'(\mu_t \phi) + \psi'((1 - \mu_t)\phi) \}$$

Sendo $\psi'(\cdot)$ a função trigama definida em (2.9), assim,

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \beta_j} \right) = -\phi \sum_{t=1}^n w_t x_{ti} x_{tj}$$

com

$$w_t = \phi \{ \psi'(\mu_t \phi) + \psi'((1 - \mu_t)\phi) \} \frac{1}{\{g'(\mu_t)\}^2}$$

Na forma matricial

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \beta^T} \right) = -\phi X^T W X,$$

em que $W = \text{diag}\{w_1, \dots, w_n\}$.

Por meio de (3.7), acha-se a derivada de 2ª ordem de $l(\beta, \phi)$ em relação a β_i e ϕ ,

$$\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \phi} = \sum_{t=1}^n \left[(y_t^* - \mu_t^*) - \phi \frac{\partial \mu_t^*}{\partial \phi} \right] \frac{1}{g'(\mu_t)} x_{ti}$$

Como $E(y_t^*) = \mu_t^*$ e $\partial \mu_t^* / \partial \phi = \text{psi}'(\mu_t \phi) \mu_t - \psi'((1 - \mu_t)\phi)(1 - \mu_t)$, com isso, tem-se

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \phi} \right) = \sum_{t=1}^n c_t \frac{1}{g'(\mu_t)} x_{ti},$$

onde $c_t = \phi \{ \psi'(\mu_t \phi) \mu_t - \psi'((1 - \mu_t)\phi)(1 - \mu_t) \}$. Em notação matricial,

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \beta_i \partial \phi} \right) = -X^T T c,$$

com $c = (c_1, \dots, c_n)^T$.

Ademais, $\partial^2 l(\beta, \phi) / \partial \phi^2$, é obtido derivando (3.9) em relação a ϕ ,

$$\begin{aligned} \frac{\partial^2 l(\beta, \phi)}{\partial \phi^2} &= \sum_{t=1}^n \frac{\partial^2 l_t(\mu_t, \phi)}{\phi^2} = \\ &= - \sum_{t=1}^n [\psi'(\mu_t, \phi) \mu_t^2 + \psi'((1 - \mu_t)\phi)(1 - \mu_t)^2 - \psi'(\phi)]. \end{aligned}$$

Sendo $d_t = \psi'(\mu_t, \phi) \mu_t^2 + \psi'((1 - \mu_t)\phi)(1 - \mu_t)^2 - \psi'(\phi)$,

$$E \left(\frac{\partial^2 l(\beta, \phi)}{\partial \phi^2} \right) = - \sum_{t=1}^n d_t = -\text{tr}(D).$$

Onde $D = \text{diag}(d_1, \dots, d_n)$ e $\text{tr}(D)$ indica o traço da matriz D .

Finalmente, a matriz de informação de Fisher é dada por

$$K = K(\beta, \phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix} \quad (3.13)$$

Onde $K_{\beta\beta} = \phi X^T W X$, $K_{\phi\phi} = \text{tr}(D)$, $K_{\beta\phi} = K_{\phi\beta}^T = X^T T c$. Nota-se que não ortogonalidade entre os parâmetro β e ϕ , indo de encontro ao que é observado nos modelos lineares generalizados (McCullagh e Nelder, 1989).

Sob condições gerais de regularidade (ver Sen e Singer, 1993, Capítulo 7), quando o tamanho da amostra for suficientemente grande,

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \sim N_{k+1} \left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, K^{-1} \right),$$

aproximadamente, no qual $\hat{\beta}$ e $\hat{\phi}$ são os estimadores de máxima verossimilhança de β e ϕ . Utilizando extensões padrões para inversas de matrizes subdivididas (ver Rao, 1973, p.29), tem-se

$$K^{-1} = K^{-1}(\beta, \phi) = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \\ K^{\phi\beta} & K^{\phi\phi} \end{pmatrix}, \quad (3.14)$$

onde

$$K^{\beta\beta} = \frac{1}{\phi} (X^T W X)^{-1} \left\{ I_k + \frac{X^T T c c^T T^T X (X^T W X)^{-1}}{\gamma \phi} \right\}$$

sendo $\gamma = \text{tr}(D) - \phi^{-1} c^T T^T X (X^T W X)^{-1} X^T T c$,

$$K^{\beta\phi} = (K^{\phi\beta})^T = \frac{-1}{\gamma \phi} (X^T W X)^{-1} X^T T c,$$

e $K^{\phi\phi} = \gamma^{-1}$. I_k é a matriz identidade de ordem k .

3.4 Estimação dos Parâmetros

Para a obtenção dos estimadores de máxima verossimilhança de β e ϕ , faz-se $U_\beta(\beta, \phi) = 0$ e $U_\phi(\beta, \phi) = 0$. Contudo, não há forma analítica para a obtenção desses estimadores, assim, métodos iterativos podem ser utilizados para solucionar esse problema. Existem vários métodos que podem ser acionados, por exemplo, Newton-Raphson, escore de Fisher, quasi-Newton, entre outros. Para este trabalho, será abordado o método de Newton-Raphson, tendo em vista a usabilidade desse método em diversas aplicações, a qual é grande.

3.4.1 Método de Newton-Raphson

Seja $\theta = (\beta^T, \phi)^T$, o vetor de parâmetros e $U(\theta) = ((U_\beta(\beta, \phi)^T, U_\phi(\beta, \phi))^T$, o vetor das funções escores com dimensão $(k+1) \times 1$. O processo iterativo de Newton-Raphson é dado pela expansão da função escore $U(\theta)$ em torno de um valor inicial $\theta^{(0)}$, ou seja

$$U(\theta) \approx U(\theta^{(0)}) + U'(\theta^{(0)})(\theta - \theta^{(0)}),$$

no qual $U'(\theta)$ é a derivada de 1ª ordem de $U(\theta)$ em relação a θ^T . Sendo $U(\theta) = 0$ e repetindo o procedimento acima, tem-se

$$\theta^{(m+1)} = \theta^{(m)} + \{-U'(\theta^{(m)})\}^{-1}U(\theta^{(m)}), \quad m = 0, 1, \dots \quad (3.15)$$

3.5 Inferências de Segunda Ordem

Para a construção de intervalos de confiança e testes de hipóteses existem algumas técnicas que podem ser aplicadas, por exemplo, o teste de Wald, teste da razão de verossimilhança, escore, entre outros. Tendo em vista que o pacote utilizado no *software* R recorre ao teste de Wald para as inferências de segunda ordem, ele será o teste abordado nesta seção.

3.5.1 Teste de Wald

O teste de Wald é um recurso para a realização de inferências assintóticas acerca dos parâmetros. Assim, a estatística de teste para $H_0 : \beta_1 = \beta_1^{(0)}$ versus $H_1 : \beta_1 \neq \beta_1^{(0)}$ é dada por

$$\omega_3 = (\hat{\beta}_1 - \beta_1^{(0)})^T (\hat{K}_{11}^{\beta\beta})^{-1} (\hat{\beta}_1 - \beta_1^{(0)}),$$

onde $\hat{K}_{11}^{\beta\beta}$ é igual a $K_{11}^{\beta\beta}$ avaliado no estimador de máxima verossimilhança irrestrito e $\hat{\beta}_1$ é o estimador de máxima verossimilhança de β_1 . Sob condições gerais de regularidade e sob H_0 , $\omega_3 \xrightarrow{d} \chi_m^2$. De modo geral, para testar a significância do i -ésimo parâmetro regressor (β_i), $i = 1, \dots, k$, utiliza-se a raiz quadrada sinalizada da estatística de Wald, ou seja, $\hat{\beta}_i / se(\hat{\beta}_i)$, em que $se(\hat{\beta}_i)$ é o erro padrão assintótico do estimador de máxima verossimilhança de β_i , obtido pela inversa da matriz de informação de Fisher avaliada nos estimadores de máxima verossimilhança. Assim, a estatística de teste sob a hipótese nula tem uma distribuição normal padrão.

Além dos testes, é interessante que seja feita a construção de intervalos de confiança para os parâmetros. Assim, intervalos de confiança para os parâmetros com $(1 - \alpha) \times 100\%$, $i = 1, \dots, k$ e $0 < \alpha < 1/2$ são dados por:

- i) $\beta_i : [\hat{\beta}_i \pm \Phi^{-1}(1 - \alpha/2)se(\hat{\beta}_i)]$
- ii) $\phi : [\hat{\phi} \pm \Phi^{-1}(1 - \alpha/2)se(\hat{\phi})]$
- iii) $exp(c\beta_i) : [exp\{c(\hat{\beta}_i \pm \Phi^{-1}(1 - \alpha/2)se(\hat{\beta}_i))\}]$
- iv) $\mu : [g^{-1}(\hat{\eta} \pm \Phi^{-1}((1 - \alpha/2)se(\hat{\eta})))]$.

Nos quais $se(\hat{\phi}) = \hat{\gamma}^{-\frac{1}{2}}$, o I.C. para $exp(c\beta_i)$ é para quando se estiver utilizando a função de ligação logito, $\hat{\eta} = x^T \hat{\beta}$ e $se(\hat{\eta}) = \sqrt{x^T \widehat{cov}(\hat{\beta})x}$. A $\widehat{cov}(\hat{\beta})$ pode ser obtida pela inversa da matriz de informação de Fisher excluindo a linha e a coluna da matriz correspondente ao parâmetro de dispersão. É necessário ressaltar que esses intervalos são válidos somente para funções de ligação estritamente crescentes.

3.6 Técnicas de Diagnóstico

Uma etapa imprescindível para a análise de regressão após o ajuste do modelo, é a de diagnóstico. Ela consiste na consistência das suposições feitas para modelo e a existência de observações discrepantes que possam interferir fortemente no ajuste do modelo.

3.6.1 Resíduo Componente do Desvio

Uma medida global de qualidade do ajuste foi proposta por Ferrari e Cribari-Neto (2004), mas originalmente colocada, em pauta, por Nelder e Wedderburn (1972) na conjuntura dos modelos lineares generalizados. Essa medida consiste na ideia de que a discrepância de um ajuste pode ser avaliada como duas vezes a diferença entre o máximo do logaritmo da verossimilhança do modelo saturado e o do modelo postulado. Dada por

$$D(y; \mu, \phi) = \sum_{t=1}^n 2\{l_t(\tilde{\mu}_t, \phi) - l_t(\mu_t, \phi)\},$$

sendo $\tilde{\mu}_t$ a solução de $\partial l_t / \partial \mu_t = 0$.

Já para avaliar se existem pontos suspeitos de serem aberrantes, utiliza-se o gráfico do resíduo padronizado com os respectivos índice de observação, sendo o resíduo padronizado dado por

$$r_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\widehat{Var}(y_t)}},$$

onde $\hat{\mu}_t = g^{-1}(x_t^T \hat{\beta})$ e $\widehat{Var}(y_t) = \{\hat{\mu}_t(1 - \hat{\mu}_t)\} / (1 + \hat{\phi})$.

3.6.2 Gráficos de Envelope Simulado Meio-Normal

A distribuição dos resíduos é desconhecida, utiliza-se a ferramenta gráficas do envelope simulado de uma meio-normal. A partir dele, pode-se observar se as respostas observadas são consistentes com o modelo e se existem tendências não-aleatórias dos resíduos dentro do envelope. Essa ferramenta é importante, pois através da análise do envelope, caso haja de fato, tendências não-aleatórias, isso pode indicar especificação incorreta da distribuição dos dados, da função de ligação ou da parte sistemática do modelo.

3.6.3 Alavancagem Generalizada

A alavancagem é medida pelos elementos da matriz de projeção e é usada para medir a importância individual de cada observação em relação ao próprio valor ajustado.

A medida proposta por Ferrari e Cribari-Neto (2004) para quantificar a alavancagem, foi avaliada em dois cenários, um quando ϕ é conhecido e outra quando o parâmetro de precisão não é conhecido. Assim, para o contexto em que ϕ é conhecido, tem-se

$$GL = TX(X^T QX)^{-1} X^T TM, \quad (3.16)$$

onde $Q = \text{diag}(q_1, \dots, q_n)$, com

$$q_t = \left\{ \phi \{ \psi'(\mu_t \phi) + \psi'((1 - \mu_t)\phi) \} + (y_t^* - \mu_t^*) \frac{g''(\mu_t)}{g'(\mu_t)} \right\} \frac{1}{\{g'(\mu_t)\}^2}, \quad t = 1, \dots, n.$$

Sendo $M = \text{diag}(m_1, \dots, m_n)$, com $m_t = 1/\{y_t(1 - y_t)\}$, $t = 1, \dots, n$.

Para uma conjuntura em que ϕ é desconhecido, tem-se

$$GL = (\beta, \phi) = GL(\beta) + \frac{1}{\gamma\phi} TX(X^T QX)^{-1} X^T T f(f^T TX(X^T QX)^{-1} X^T TM - b^T),$$

onde $f = (f_1, \dots, f_n)$ com $f_t = \{c_t - (y_t^* - \mu_t^*)\}$ e $b = (b_1, \dots, b_n)$, sendo $b_t = -(y_t - \mu_t)/\{y_t(1 - y_t)\}$, $t = 1, \dots, n$.

Nota-se que, quando ϕ é suficientemente grande, $GL(\beta, \phi) \approx GL(\beta)$.

3.6.4 Influência

Outra etapa extremamente importante do diagnóstico do modelo, é a parte de avaliação de influência, a qual essencialmente consiste em detectar observações que influenciem demais nas estimativas dos parâmetros. Para isso, utiliza-se a aproximação usual para distância

de Cook (Cook, 1997), dada por

$$C_t = \frac{h_{tt} r_t^2}{k(1 - h_{tt})^2},$$

Essa medida combina alavancagem e resíduos. Logo, C_t será grande quando r_t for grande ou quando h_{tt} for próximo de um.

Por fim, utiliza-se o gráfico de C_t contra a ordem das observações para avaliar a existência de pontos suspeitos de serem influentes.

4 APLICAÇÃO

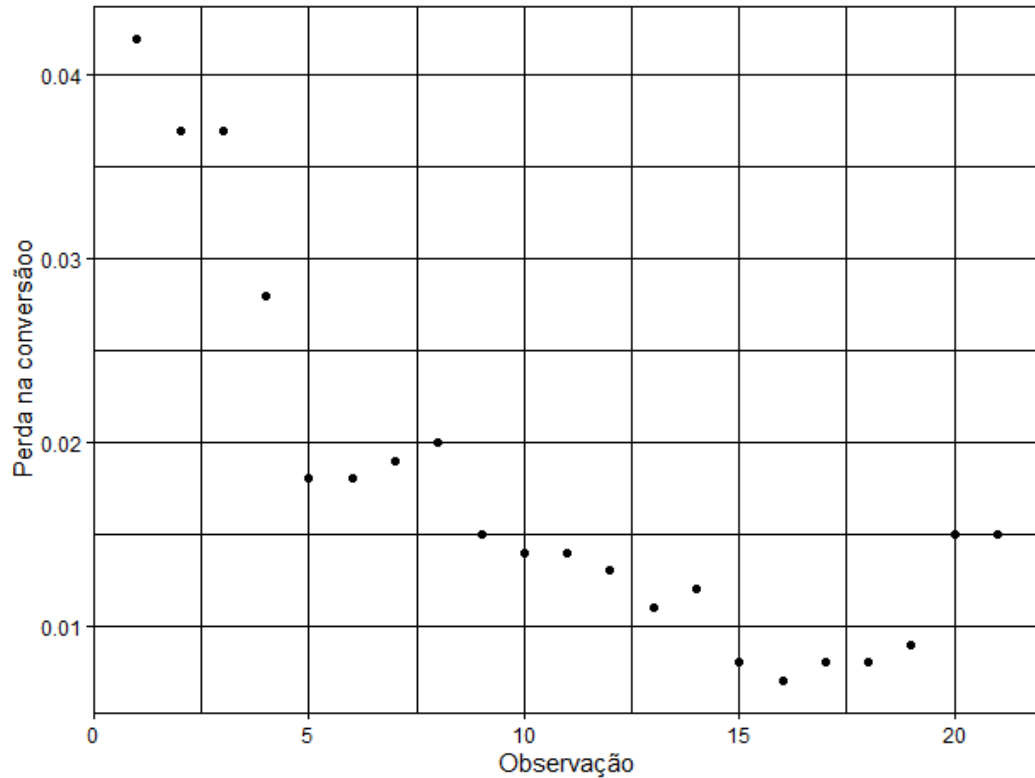
Tendo em vista a necessidade de avaliar o modelo numa situação prática após falar sobre a teoria que embasa a utilização do modelo, utilizou-se um exemplo de Brownlee (1965, p.64).

Os dados da Tabela 1 foram obtidos em 21 dias de processos de oxidação de amônia como um estágio para a produção de ácido nítrico em uma planta industrial. A variável resposta é proporção de amônia que não foi convertida em ácido nítrico. As variáveis explicativas são: corrente de ar, temperatura da água e concentração de ácido ($10 \times \text{concentrao decido} - 50$).

Tabela 1 – Dados do estudo de oxidação da Amônia

Observacao	Perda na conversão (y)	Corrente de ar (x_1)	Temperatura da água (x_2)	Concentração de ácido (x_3)
1	0.042	80	27	89
2	0.037	80	27	88
3	0.037	75	25	90
4	0.028	62	24	87
5	0.018	62	22	87
6	0.018	62	23	87
7	0.019	62	24	93
8	0.020	62	24	93
9	0.015	58	23	87
10	0.014	58	18	80
11	0.014	58	18	89
12	0.013	58	17	88
13	0.011	58	18	82
14	0.012	58	19	93
15	0.008	50	18	89
16	0.007	50	18	86
17	0.008	50	18	72
18	0.008	50	19	79
19	0.009	50	20	80
20	0.015	56	20	82
21	0.015	70	20	91

Figura 3 – Gráfico da perda de conversão x índice das observações.



Na Figura 3 foi analisada a perda de conversão para cada observação, de acordo com Atkinson (1985, p.130), ele argumenta que como a variável resposta é não-negativa, é possível ajustar um modelo linear fazendo uma transformação conveniente na perda da conversão. A proposta em questão é ajustar um modelo utilizando a modelagem beta, assim, sem a necessidade de fazer transformações nas variáveis.

Para realizar o ajuste do modelo, foi utilizado o pacote *betareg*, de autoria de Alexandre B. Simas e Andrea V. Rocha. Utilizando o modelo dado por

$$g(\mu_t) = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3}, \quad (4.1)$$

em que $g(\cdot)$ representa a função logito.

Fazendo o ajuste do modelo, tem-se

Tabela 2 – Estimativas dos parâmetros		
Parâmetro	Estimativa	Valor-p
β_0	-7,590543	2×10^{-16}
β_1	0,029070	$2,87 \times 10^{-6}$
β_2	0,075362	$5,83 \times 10^{-5}$
β_3	0,001188	0,89
ϕ	2413,6	-

Nota-se que pelo teste de significância para β_1 , o valor-p foi superior ao nível de significância utilizado pelo pacote *betareg*, sendo assim, estatisticamente não significativo, logo, pode-se ajustar o modelo somente com x_1 e x_2 . Assim, as novas estimativas dos parâmetros obtidas foram

Tabela 3 – Estimativas dos parâmetros sem a variável x_3 .

Parâmetro	Estimativa	Valor-p
β_0	-7,503	2×10^{-16}
β_1	0,029	$1,79 \times 10^{-6}$
β_2	0,0754	$5,10 \times 10^{-5}$
ϕ	2410,7	-

Nota-se que todos os níveis descritivos estão próximos de zero, logo, são estatisticamente significativos. Utilizando o pseudo R_p^2 , tem-se que $R_p^2 = 0,9041$, essa métrica de qualidade do ajuste está fornecendo a proporção da variabilidade da variável resposta que é explicada pela variável explicativa, sendo ela dada pelo quadrado do coeficiente de correlação amostral entre $g(y)$ e $\hat{\eta}$. Logo, está indicando um bom ajuste.

Realizando o diagnóstico do modelo, tem-se

Figura 4 – Gráficos para o diagnóstico do modelo.

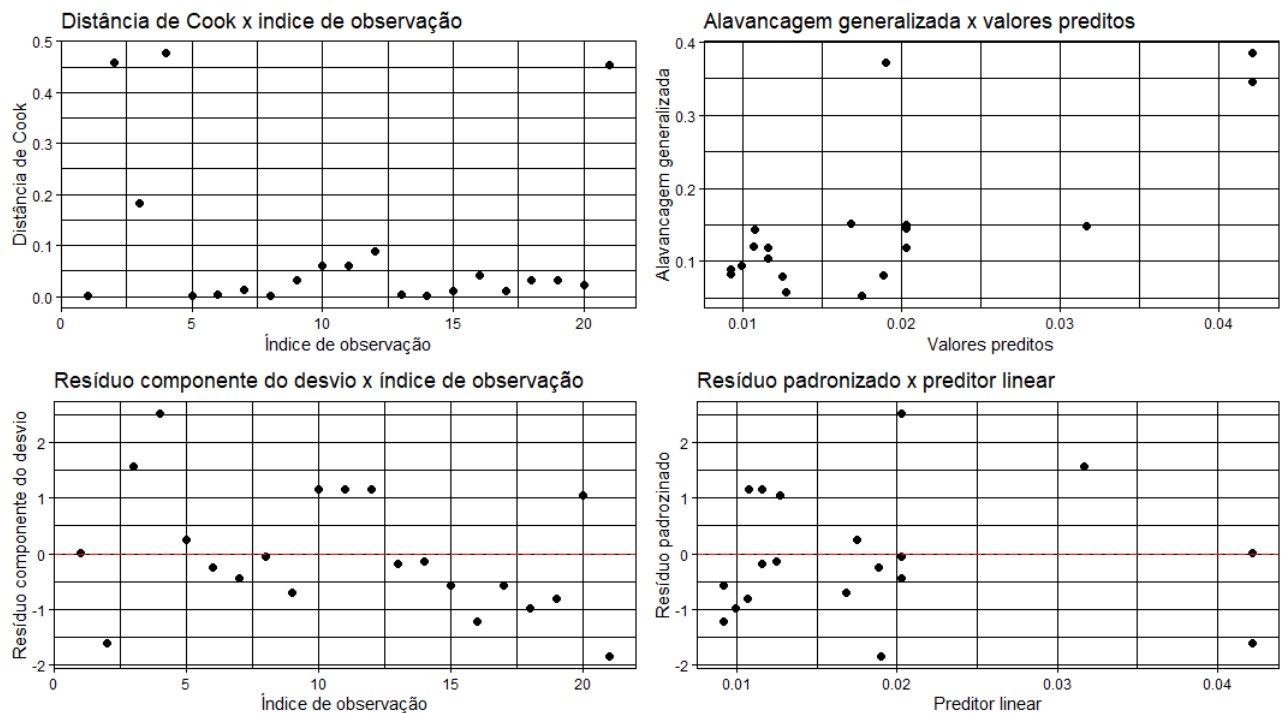
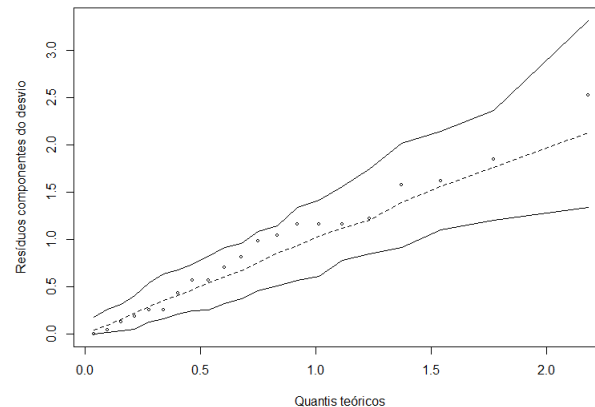


Figura 5 – Envelope simulado meio-normal para os resíduos componentes do desvio.



Analisando as Figuras 4, nota-se que o maior resíduo componente do desvio, em valor absoluto, corresponde à observação 4, pelo gráfico dos resíduos componentes do desvio x índice de observação. Ademais, observa-se que pelo gráfico do resíduo padronizado x preditor linear, os pontos estão distribuídos de forma aleatória, não apresentando nenhuma tendência. Também, é notório que todos os pontos se encontram dentro do envelope simulado, pela Figura 5.

À luz dessas considerações, conclui-se que o modelo proposto parece útil para modelar os dados presentes na Tabela 1. Ademais, pode-se fazer também ajustes para o mesmo modelo sem utilizar as observações que acabaram se destacando, visando avaliar a influência delas na estimação dos parâmetros, por exemplo. No mais, foi possível observar como ocorre o ajuste de um modelo de regressão beta.

5 CONCLUSÃO

Nesse trabalho foi apresentado o modelo de regressão beta, proposto por Ferrari e Cribari-Neto (2004). Foram apresentadas as características da distribuição beta, desde dos momentos, da curtose, até a estimação dos parâmetros pelo método de máxima verossimilhança.

Posteriormente, abordou-se o modelo propriamente dito, sendo colocadas, em pauta, as funções de ligação, a função escore, a matriz de informação de Fisher, a estimação dos parâmetros, a inferência de segunda ordem e as técnicas de diagnóstico.

Ademais, foi mostrado um exemplo prático de ajuste, relacionado à oxidação da Amônia. Nessa aplicação foi possível compreender a proposta do modelo em questão, a qual consiste em modelar variáveis resposta num intervalo unitário (0,1).

À luz dessas considerações, constata-se que o modelo de regressão beta é uma proposta bastante interessante e viável para modelar taxas e proporções.

REFERÊNCIAS

- Cribari-Neto F, Lima LB (2007). “**A Misspecification Test for Beta Regressions.**” *Technical report*.
- Cribari-Neto F, Zeileis A (2010). “**Beta Regression in R.**” *Journal of Statistical Software*, 34(2), 1–24. URL <http://www.jstatsoft.org/v34/i02/>.
- Espinheira PL, Ferrari SLP, Cribari-Neto F (2008a). “**Influence Diagnostics in Beta Regression.**” *Computational Statistics & Data Analysis*, 52(9), 4417–4431.
- Espinheira PL, Ferrari SLP, Cribari-Neto F (2008b). “**On Beta Regression Residuals.**” *Journal of Applied Statistics*, 35(4), 407–419.
- Ferrari SLP, Cribari-Neto F (2004). “**Beta Regression for Modelling Rates and Proportions.**” *Journal of Applied Statistics*, 31(7), 799–815.
- Oliveiras, M. S. (2004). “**Um Modelo de Regressão Beta: Teoria e Aplicações**”.

APÊNDICE A – INFORMAÇÃO DE FISHER

Para obter a matriz de informação de Fisher dos parâmetros p e q , utiliza-se o logaritmo natural da função de verossimilhança para uma amostra y_1, \dots, y_n , sendo dado por

$$l(p, q) = -n \log B(p, q) + (p-1) \sum_{i=1}^n \log y_i + (q-1) \sum_{i=1}^n \log(1-y_i).$$

Substituindo $B(p, q)$ dado em (2.5), na expressão acima, tem-se

$$l(p, q) = -n[\log \Gamma(p) + \log \Gamma(q) - \log \Gamma(p+q)] + (p-1) \sum_{i=1}^n \log y_i + (q-1) \sum_{i=1}^n \log(1-y_i).$$

Já as derivas de primeira ordem de $l(p, q)$ em relação à p e à q , são dadas por

$$\frac{\partial l(p, q)}{\partial p} = -n[\psi(p) - \psi(p+q)] + \sum_{i=1}^n \log y_i$$

e

$$\frac{\partial l(p, q)}{\partial q} = -n[\psi(q) - \psi(p+q)] + \sum_{i=1}^n \log(1-y_i).$$

As derivadas de segundas ordem são

$$\frac{\partial^2 l(p, q)}{\partial p^2} = -n[\psi'(p) - \psi'(p+q)] \quad (\text{A.1})$$

e

$$\frac{\partial^2 l(p, q)}{\partial q^2} = -n[\psi'(q) - \psi'(p+q)]. \quad (\text{A.2})$$

Ademais, a derivada cruzada de segunda ordem é

$$\frac{\partial^2 l(p, q)}{\partial p \partial q} = n\psi'(p+q). \quad (\text{A.3})$$

Assim, a matriz de informação de Fisher é dada por

$$K(p, q) = -n \begin{pmatrix} \psi'(p) - \psi'(p+q) & -\psi'(p+q) \\ -\psi'(p+q) & \psi'(q) - \psi'(p+q) \end{pmatrix}.$$