Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Estatística

Gisele de Oliveira Maia

Modelo Binomial Semiparamétrico

Juiz de Fora
2017

Gisele de Oliveira Maia

Modelo Binomial Semiparamétrico

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do título de Bacharel em Estatística.

Orientador: Clécio da Silva Ferreira

Juiz de Fora

2017

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

de Oliveira Maia, Gisele.
Modelo Binomial Semiparamétrico / Gisele de Oliveira Maia. – 2017.
53 f. : il.

Orientador: Clécio da Silva Ferreira
Monografia – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Departamento de Estatística, 2017.

1. Semiparamétrico. 2. Curva Suave. 3. Modelo Binomial Semiparamétrico. I. da Silva Ferreira, Clécio, orient. II. Título.

Gisele de Oliveira Maia

Modelo Binomial Semiparamétrico

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do título de Bacharel em Estatística.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Clécio da Silva Ferreira - Orientador
Universidade Federal de Juiz de Fora

Professora Dra. Camila Borelli Zeller
Universidade Federal de Juiz de Fora

Professor Dr. Ronaldo Rocha Bastos
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Agradecimento em especial aos meus pais e minhas irmãs.

Meu orientador, Clécio, por todos ensinamentos e dedicação.

A todos meus amigos, familiares e professores do Departamento de Estatística.

A todos que contruibuíram de alguma forma para realização deste trabalho, meu muito obrigada.

RESUMO

Em diversos estudos, variáveis de interesse podem apresentar relações lineares e não lineares com variáveis auxiliares. Por isso a importância de se trabalhar com os modelos semiparamétricos, onde tanto estimamos os parâmetros para a parte paramétrica quanto para a parte não paramétrica, esta última sendo estimada através de uma curva suave proposta por Green e Silverman[7] (1994) e Eilers e Marx[4] (1996). Neste trabalho, assumimos que a variável resposta segue uma distribuição Binomial, onde a componente sistemática terá duas composições, uma linear e uma não linear. Os parâmetros do modelo são estimados através do método de Newton-Raphson, com o auxílio da função escore e da matriz de informação de Fisher. Os cálculos foram realizados considerando as três funções de ligação, a saber *logit*, *probit* e complemento log-log. O parâmetro de suavização é obtido através da minimização da função de validação cruzada. Foram realizadas simulações e aplicações a dados reais utilizando o *software* R Core Team[11] (2017) com auxílio de algoritmos feitos neste trabalho.

Palavras-chave: Semiparamétrico. Curva Suave. Modelo Binomial Semiparamétrico.

ABSTRACT

In several studies of interest variables may present linear and non-linear relationships with auxiliary variables. Therefore, the importance of working with semi-parametric models, where we both estimate the parameters for the parametric part and the non-parametric, the latter being estimated through a smooth curve proposed by Green and Silverman[7] (1994) and Eilers and Marx[4] (1996). In this work we assume that the response variable follows a Binomial distribution where the systematic component will have two compositions, one linear and one nonlinear. The parameters of the model are estimated using the Newton-Raphson method, with the aid of the score function and Fisher's information matrix. The calculations were performed considering the three link functions, namely logit, probit and complement log-log. The smoothing parameter is obtained by minimizing the cross-validation function. Simulations and real data applications were performed using the R Core Team software[11] (2017) with the aid of algorithms made in this work that enabled the validation of the proposed model to be validated.

Key-words: Semiparametric. Smooth Curve. Semiparametric Binomial Model.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Os modelos clássicos de regressão paramétrica são amplamente utilizados quando há interesse em estudar a relação entre uma variável $\mathbf{Y}$, chamada variável resposta ou variável dependente, com outras variáveis independentes, chamadas de variáveis explicativas. Os efeitos das variáveis explicativas sobre a variável resposta são lineares e aditivos. O modelo clássico de regressão paramétrica pode ser assim definido $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, onde o objetivo inicial é estimar o vetor $\boldsymbol{\beta}$. Estes parâmetros possuem uma interpretação física que possibilitam explicar a relação entre as variáveis independentes com a variável dependente e podemos estima-los pelo método de mínimos quadrados e máxima verossimilhança, este último supondo que os erros são não correlacionados e seguem uma distribuição normal com média zero e variância σ^2, assim $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. Para mais informações sobre regressão paramétrica ver Charnet et al.[1] (1999).

Podemos ter variáveis explicativas que possuem efeito não linear sobre a variável resposta, estas variáveis são denominadas de variáveis não paramétricas. Geralmente, não possuímos conhecimento a priori a respeito da forma desta variável. Desta relação não paramétrica entre a variável explicativa e a variável resposta temos uma curva suave. Então, na regressão não paramétrica o principal objetivo é a estimação desta curva suave, utilizando conceitos de splines cúbicos naturais. Aqui, não temos parâmetros interpretáveis, somente a curva estimada. Toda a teoria sobre este assunto pode ser aprofundada em Green e Silverman[7] (1994) e Eubank[5] (1999).

A junção em um mesmo modelo de variáveis explicativas paramétricas e não paramétricas dá origem ao modelo semiparamétrico, que é denotado da seguinte forma $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{g} + \epsilon$, em que $\mathbf{X}$ é a matriz composta pelas variáveis explicativas e $\mathbf{g}$ a variável não paramétrica. Neste trabalho, assumimos também que os erros são não correlacionados, seguindo uma distribuição normal com média zero e variância constante, temos que $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{g}, \sigma^2\mathbf{I})$. Além disso, utilizamos dois métodos para se trabalhar com a variável não paramétrica, e que se diferenciam na construção das matrizes $\mathbf{N}$ e $\mathbf{K}$(matriz referente ao termo da penalty que será introduzido no capítulo 1). Estes dois métodos são: o método de Green e Silverman[7] (1994) e Eilers e Marx[4] (1996).

O próximo passo no trabalho foi aplicar regressão semiparamétrica em modelos lineares generalizados, especificamente, em modelo binomial, ou seja, trabalhamos com modelo binomial semiparamétrico. Embora Green e Silverman[7] e Hastie e Tibshirani [8] (1987) apresentem o modelo, os mesmos não apresentam os detalhes de estimações dos parâmetros e dos cálculos dos erros-padrão.

Para estimação de $\boldsymbol{\beta}$ e $\mathbf{g}$ utilizamos o método de Newton-Raphson, que necessita da matriz de informação esperada de Fisher e da função escore, que serão detalhados neste trabalho. Este algoritmo foi implementado no software R Core Team[11] (2017) e será

usado no estudo de simulação e aplicação a dados reais.

2 MODELO NÃO PARAMÉTRICO

Antes de introduzirmos os modelos semiparamétricos, iremos focar neste capítulo nos modelos não paramétricos, onde iniciaremos a apresentação de um dos métodos para se trabalhar com a variável não paramétrica. Todo este capítulo foi baseado nos capítulos 1, 2 e 3 do livro de Green e Silverman[7] (1994).

Suponha o seguinte modelo:

$$\mathbf{y} = \mathbf{g} + \boldsymbol{\epsilon},$$

em que $\mathbf{y}$ é a variável resposta do modelo, $\mathbf{g}$ uma curva qualquer e $\boldsymbol{\epsilon}$ o erro.

Como $\mathbf{g}$ é uma curva não podemos utilizar regressão linear para ajustar este modelo. Na regressão linear, o principal objetivo é estimar os parâmetros do modelo e assim obter as interpretações físicas entre as variáveis explicativas com a variável resposta, mas no modelo não paramétrico, estamos interessados em estimar a curva $\mathbf{g}$. Assim, não temos mais parâmetros interpretáveis.

Na estimação da curva $\mathbf{g}$ trabalhamos com os splines, que são curvas estimadas com o auxílio de pontos que serão denominados mais adiante de nós, onde os coeficientes dos polinômios relacionados a curva são estimados por mínimos quadrados ou máxima verossimilhança.

2.1 PENALIDADE DE RUGOSIDADE

Ao utilizarmos o método de mínimos quadrados para estimação da curva $\mathbf{g}$, temos que acrescentar um termo à equação de mínimos quadrados. Este termo está relacionado a penalidade de rugosidade da curva e temos que quantificá-lo. Está penalidade estará acrescentando informações da variável não paramétrica $\mathbf{g}$.

Uma maneira de definir a rugosidade de uma curva $\mathbf{g}$ diferenciável definida no intervalo [a,b] é calcular a integral da derivada segunda de $\mathbf{g}$ no intervalo [a,b], $\int_a^b \{g^{(2)}(t)\}^2 dt$. Basicamente, a integral estará medindo o comprimento da curva $\mathbf{g}$.

Vamos considerar uma curva $\mathbf{g}$ diferenciável duas vezes, definida em [a,b] e um parâmetro de suavização $\alpha > 0$. Definimos a soma de quadrados penalizados da seguinte maneira

$$S(\mathbf{g}) = \sum_{i=1}^{n} \{Y_i - g(t_i)\}^2 + \alpha \int_a^b \{\mathbf{g}^{(2)}(t)\}^2 dt \tag{2.1}$$

O estimador $\hat{\mathbf{g}}$ de mínimos quadrados penalizados é definido como o argumento mínimo da função S($\mathbf{g}$).

O valor de α tem o compromisso de atender tanto a suavidade como também a qualidade do ajuste. Se o valor de α tende ao infinito, S($\mathbf{g}$) vai estar dando maior peso para o termo $\alpha \int_a^b \{g^{(2)}(t)\}^2 dt$, assim a curva apresentará pouca curvatura, como mostrado no exemplo da Figura 1, mas se α for pequeno o termo $\sum_{i=1}^{n} \{Y_i - g(t_i)\}^2$ terá maior peso em S($\mathbf{g}$), o que fará a curva acompanhar todas as observações, Figura 2.

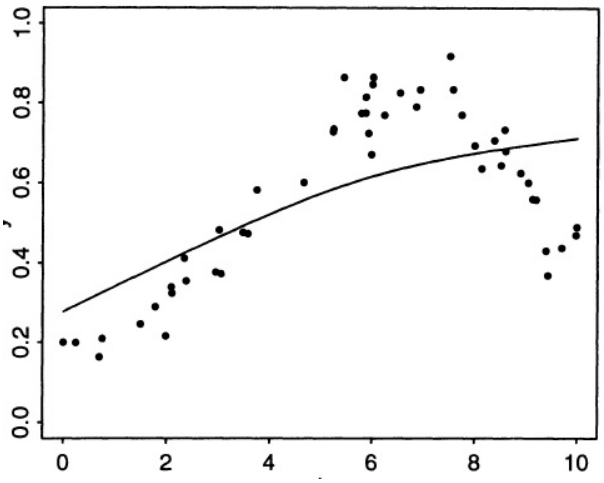


Figura 1 – Exemplo de curva estimada $\hat{\mathbf{g}}$ com alto valor de α (Fonte: Green and Silverman-1994)

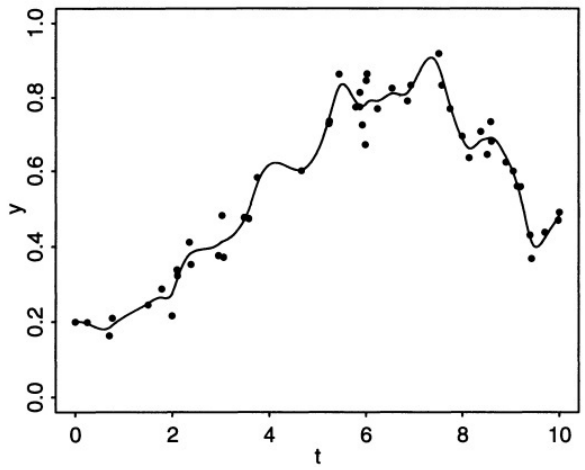


Figura 2 – Exemplo de curva estimada $\hat{\mathbf{g}}$ com pequeno valor de α (Fonte: Green and Silverman-1994)

2.2 SPLINE CÚBICO

Nesta seção, iremos explicar como obtemos a estimação da curva $\mathbf{g}$ que minimiza a soma de quadrados penalizados. Esta estimação envolve os splines cúbicos.

Suponha que sejam dados os números reais $t_1, t_2, ..., t_n$ pertencentes ao intervalo [a,b], ou seja, a$<t_1<t_2<...<t_n<$b. $\mathbf{g}$ definida no intervalo [a,b] será um spline cúbico se as duas condições forem satisfeitas:

i) Em cada intervalo (a,t_1),(t_1,t_2),(t_2,t_3),...,(t_n,b) g é um polinômio cúbico.

ii) As peças polinomiais se encaixam nos pontos t_i de forma que a primeira e segunda derivada de **g** são contínuas em cada t_i e, portanto, em todo o intervalo [a,b].

Os pontos t_i são denominados de nós.

Uma maneira de especificar um spline cúbico é dar os quatro coeficientes polinomiais de cada peça cúbica, ficando da seguinte forma

$$g(t) = d_i(t - t_i)^3 + c_i(t - t_i)^2 + b_i(t - t_i) + a_i, t_i < t < t_{i+1}$$

a_i, b_i, c_i, d_i são constantes, $i = 0, 1, ..., n$ e definimos $t_0 = a$ e $t_{n+1} = b$.

Um spline cúbico em um intervalo [a,b] será dito um spline cúbico natural (SCN) se suas derivadas segunda e terceira forem iguais a zero em a e b. Isso implica que $d_0 = c_0 = d_n = c_n = 0$, de modo que **g** é linear nos dois intervalos extremos [a,t_1] e [t_n,b].

Há uma representação mais conveniente matematicamente e computacionalmente do spline cúbico natural, dando seu valor e derivada segunda em cada um dos nós. Esta representação é chamada de Representação Via Valor da Derivada Segunda. Seja **g** uma SCN com nós $t_1 < ... < t_n$. Definimos

$$g_i = g(t_i) \quad e \quad \gamma_i = g^{(2)}(t_i), \quad i = 1, 2, ...n.$$

Assim, temos o vetor **g**, $(g_1, g_2, ..., g_n)^T$ e o vetor $\boldsymbol{\gamma}$, $(\gamma_2, \gamma_3, ..., \gamma_{n-1})^T$. Estes dois vetores especificam completamente a curva **g**, e é possível dar fórmulas explícitas em termos de **g** e $\boldsymbol{\gamma}$ para o valor e derivadas de **g** em qualquer ponto t. Com isso **g** pode ser plotado para qualquer grau de precisão desejado.

Pode ocorrer que nem todos os vetores possíveis de **g** e $\boldsymbol{\gamma}$ representem um spline cúbico natural, para isso eles devem seguir uma condição necessária e suficiente para que os vetores representem um spline cúbico natural na sequência de nós dados. Essa condição depende das duas matrizes **Q** e **R** que serão determinadas a seguir. Temos

$$h_i = t_{i+1} - t_i, i = 1, ..., n - 1$$

Q é definida como uma matriz com os elementos q_{ij} de dimensão nx(n-2), $i = 1, ..., n$ e $j = 2, ..., n - 1$, dado por

$$q_{j-1,j} = h_{j-1}^{-1}, \quad q_{jj} = -h_{j-1}^{-1} - h_j^{-1}, \quad q_{j+1,j} = h_j^{-1},$$

em que $q_{ij} = 0$ para $|i - j| \geq 2$.

As colunas de $\mathbf{Q}$ são enumeradas de tal forma que começam em j=2, assim o primeiro elemento da matriz $\mathbf{Q}$ é q_{12}.

A matriz simétrica $\mathbf{R}$ de dimensão (n-2)x(n-2) com os elementos $r_{ij}, i = 2, ..n-1$ e $j = 2, ..., n-1$, é assim definida

$$r_{ii} = \frac{1}{3}(h_{i-1} + h_i), \quad i = 2, ..., n-1$$
$$r_{i,i+1} = \frac{1}{6}h_i, \quad i = 2, ..., n-2,$$

em que $r_{ij} = 0$ para $|i-j| \geq 2$.

A matriz $\mathbf{R}$ é diagonal dominante e estritamente positiva definida. Assim, podemos definir uma matriz $\mathbf{K}$ por

$$\mathbf{K} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^{\mathbf{T}}$$

Para os vetores $\mathbf{g}$ e $\boldsymbol{\gamma}$ especificarem um spline cúbico natural $\mathbf{g}$ a condição a seguir deve ser satisfeita

$$\mathbf{Q}^{\mathbf{T}}\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} \tag{2.2}$$

Se a condição (2.2) estiver satisfeita a penalidade de rugosidade irá satisfazer

$$\int_a^b \mathbf{g}^{(2)}(t)^2 dt = \boldsymbol{\gamma}^T\mathbf{R}\boldsymbol{\gamma} = \mathbf{g}^{\mathbf{T}}\mathbf{K}\mathbf{g} \tag{2.3}$$

Satisfeita a condição (2.2) e consequentemente a (2.3), podemos reescrever a soma de quadrados penalizados (2.1) da seguinte forma matricialmente

$$S(\mathbf{g}) = (\mathbf{Y} - \mathbf{g})^T(\mathbf{Y} - \mathbf{g}) + \alpha\mathbf{g}^T\mathbf{K}\mathbf{g}$$

Sendo $\mathbf{Y} = (Y_1, ..., Y_n)^T$, temos que

$\sum_{i=1}^n \{Y_i - g(t_i)\}^2 = (\mathbf{Y} - \mathbf{g})^T\ (\mathbf{Y} - \mathbf{g})$

2.3 EXISTÊNCIA E UNICIDADE DO MINIMIZADOR DE UMA CURVA SPLINE

Suponha que $\mathbf{g}$ é um spline cúbico natural, com vetores $\mathbf{g}$ e $\boldsymbol{\gamma}$ e matrizes $\mathbf{Q}$ e $\mathbf{R}$. Assim, temos que

$$\begin{aligned} S(\mathbf{g}) &= (\mathbf{Y} - \mathbf{g})^T(\mathbf{Y} - \mathbf{g}) + \alpha\mathbf{g}^T\mathbf{K}\mathbf{g} \\ &= \mathbf{g}^T(\mathbf{I} + \alpha\mathbf{K})\mathbf{g} - \mathbf{Y}^T\mathbf{g} + \mathbf{Y}^T\mathbf{Y} - \mathbf{g}^T\mathbf{Y} \end{aligned} \tag{2.4}$$

Uma vez que $\alpha\mathbf{K}$ é não-negativa definida, a matriz $\mathbf{I} + \alpha\mathbf{K}$ é estritamente positiva definida. Portanto, temos que (2.4) possui um mínimo único dado por

$$\mathbf{g} = (\mathbf{I} + \alpha\mathbf{K})^{-\mathbf{1}}\mathbf{Y} \tag{2.5}$$

O vetor $\mathbf{g}$ define o spline $\mathbf{g}$ de forma exclusiva. Assim, no espaço de todos os splines cúbicos naturais com nós nos pontos t_i, S($\mathbf{g}$) tem um mínimo exclusivo dado por (2.5).

Para finalizar a seção, considere, n ≥ 3 e os pontos $t_1, ..., t_n$ que satisfazem $a < t_1 < ... < t_n < b$. Os pontos de dados $Y_1, ..., Y_n$, α um parâmetro de suavização estritamente positivo, um spline cúbico natural $\hat{\mathbf{g}}$ com nós $t_1, ..., t_n$ e uma qualquer $\mathbf{g}$ em $S_2[a, b]$, temos

$$\mathrm{S}(\hat{\mathbf{g}}) \leq \mathrm{S}(\mathbf{g})$$

com igualdade somente se $\mathbf{g}$ e $\hat{\mathbf{g}}$ são idênticos.

2.4 ESCOLHA DO PARÂMETRO DE SUAVIZAÇÃO α

Existem duas abordagens diferentes sobre a questão de escolher o parâmetro de suavização. A primeira é considerar a escolha livre do parâmetro de suavização como uma característica vantajosa do procedimento, ou seja, é uma abordagem mais subjetiva. Já a outra abordagem sugere que existe uma necessidade de um método automático pelo qual o valor do parâmetro de suavização seja escolhido pelos dados.

Temos vários procedimentos automáticos diferentes, mas neste trabalho vamos abordar a validação cruzada, que é o mais conhecido e o que foi utilizado em nosso estudo.

2.4.1 VALIDAÇÃO CRUZADA

A idéia básica da validação cruzada é em termos de previsão. Supondo que o erro aleatório tenha uma média zero, a curva de regressão real $\mathbf{g}$ tem a propriedade de que, se uma observação Y for tomada em um ponto t, o valor g(t) é o melhor preditor de Y em termos de erro quadrático médio. Assim, uma boa escolha do estimador g(t) seria aquele que deu um pequeno valor de $\{\mathrm{Y} - \hat{\mathrm{g}}\ (\mathrm{t})\}^2$ para uma nova observação Y no ponto t.

Tendo um valor fixo do parâmetro de suavização α, consideremos uma observação Y_i em t_i como sendo uma nova observação, omitida a partir do conjunto de dados utilizados para estimar a própria curva. Denote $\hat{g}^{(-i)}(t, \alpha)$ a curva estimada a partir dos dados restantes, ou seja, $\hat{g}^{(-i)}(t, \alpha)$ é o minimizador de (2.1).

A eficácia do procedimento com o parâmetro de suavização α pode ser quantificada pela função escore da validação cruzada

$$CV(\alpha) = n^{-1} \sum_{i=1}^{n} \left\{ Y_i - \hat{g}^{(-i)}(t_i, \alpha) \right\}^2$$

O objetivo da validação cruzada dada acima é escolher um valor de α que minimize a função CV(α). Como a função CV(α) não garante um mínimo exclusivo, temos o cálculo de CV(α) para uma série de valores de α, deve-se tomar cuidado na escolha do método para o cálculo do CV(α).

Podemos reescrever a função CV(α) de uma forma mais simples da seguinte maneira

$\mathbf{g} = (\mathbf{I} + \alpha\mathbf{K})^{-1}\mathbf{Y} = \mathbf{A}(\alpha)\mathbf{Y}$, então temos que

$\mathbf{A}(\alpha) = \mathbf{(I + \alpha\ K)^{-1} = (I + \alpha\ QR^{-1}Q^T)^{-1}}$

A matriz $\mathbf{A(\alpha)}$, denominada matriz chapéu (*hat*), mapeia o vetor de valores observados Y_i para seus valores previstos $\hat{g}(t_i)$ ou $\hat{Y}_i$.

Assim, podemos reescrever CV(α)

$$CV(\alpha) = n^{-1} \sum_{i=1}^{n} \left(\frac{Y_i - \hat{g}(t_i)}{1 - A_{ii}} \right)^2$$

em que $\hat{\mathbf{g}}$ é o spline suave calculado a partir dos dados (t_i, Y_i) com o parâmetro de suavização α.

Temos também o escore da validação cruzada generalizada, definida por Golub, Heath e Wahba[6] (1979), dado por

$$GCV(\alpha) = n^{-1} \frac{\sum_{i=1}^{n} \{Y_l - \hat{g}(t_i)\}^2}{\{1 - n^{-1} tr(A(\alpha))\}^2}$$

3 MODELO SEMIPARAMÉTRICO

Neste capítulo introduzimos o modelo semiparamétrico normal, pois no início de nossos estudos, na avaliação dos métodos para se trabalhar com a variável não paramétrica e alguns conceitos que serão necessários no capítulo seguinte, optamos por começar com a variável resposta seguindo a distribuição normal.

Considere o modelo

$$Y_i = \mathbf{x_i^T}\boldsymbol{\beta} + g(t_i) + \epsilon_i \tag{3.1}$$

em que Y_i representa cada observação e temos as variáveis explicativas: um vetor $\mathbf{x}_i$, que representada cada observação das p variáveis paramétricas e uma curva $\mathbf{g}$. O vetor $\boldsymbol{\beta}$ de coeficientes da regressão e uma curva suave $\mathbf{g}$ serão estimados. ϵ_i é o erro referente a cada observação.

Obs: Em nossos estudos optamos por não incluir o intercepto, pois poderia haver um problema de identificabilidade na parte não paramétrica. Assim a matriz $\mathbf{X}$ composta pelas variáveis explicativas paramétricas não terá a coluna formada por 1's.

O modelo representado em (3.1) é denominado modelo semiparamétrico, pois Y depende de certas variáveis explicativas de forma linear (variáveis paramétricas), que é o caso das variáveis explicativas $\mathbf{x}$ e em outras depende de forma não-linear (variável não paramétrica) que é a variável $\mathbf{g}$.

3.1 MÍNIMOS QUADRADOS PENALIZADOS PARA MODELOS SEMIPARAMÉTRICOS

Podemos estimar $\boldsymbol{\beta}$ e $\mathbf{g}$ pelo método de mínimos quadrados, minimizando a quantidade

$$\sum_{i=1}^{n}\{Y_i - \mathbf{x_i}^T\boldsymbol{\beta} - g(t_i)\}^2.$$

Mas está abordagem falhará, pois não temos restrições para a estimação da curva $\mathbf{g}$. Assim devemos usar o método dos mínimos quadrados penalizados, onde acrescentamos o termo de penalização

$$\sum_{i=1}^{n}\{Y_i - \mathbf{x_i}^T\boldsymbol{\beta} - g(t_i)\}^2 + \alpha\int g^{(2)}(t)^2dt. \tag{3.2}$$

3.2 MATRIZ DE INCIDÊNCIA

Denote por $s_1, s_2, ..., s_q$ os valores ordenados e distintos do intervalo $t_1, t_2, ..., t_n$. A relação entre $t_1, t_2, ..., t_n$ e s_1, s_2,....,s_q é capturada pela matriz de incidência $\mathbf{N}$, com entradas $N_{ij} = 1$ se $t_i = s_j$ e 0 caso contrário. A matriz $\mathbf{N}$ tem dimensão n x q.

Reescrevendo (3.2) na forma matricial, temos

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ng})^{\mathbf{T}}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ng}) + \alpha\mathbf{g}^{\mathbf{T}}\mathbf{Kg} \tag{3.3}$$

A função em 3.3 é minimizada quando $\boldsymbol{\beta}$ e $\mathbf{g}$ satisfazem a equação

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{N} \\ \mathbf{N}^T\mathbf{X} & \mathbf{N}^T\mathbf{X} + \alpha\mathbf{K} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{g} \end{pmatrix} = \begin{bmatrix} \mathbf{X}^T \\ \mathbf{N}^T \end{bmatrix} \mathbf{Y}$$

3.3 MÁXIMA VEROSSIMILHANÇA PENALIZADA PARA MODELOS SEMIPARAMÉTRICOS

Primeiro, considere n observações independentes Y_1,...,Y_n e o vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{g}, \sigma^2)$, iremos supor que o erro no modelo (3.1) segue a seguinte distribuição: $\epsilon \sim N(0, \sigma^2\mathbf{I})$, assim $\mathbf{Y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Ng}, \sigma^2\mathbf{I})$.

Temos que,

$$f(\mathbf{y}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ng})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ng})\}.$$

Assim, podemos escrever o logaritmo da função de máxima verossimilhança

$$l(\boldsymbol{\theta}) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ng})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ng}) \tag{3.4}$$

O logaritmo da máxima verossimilhança penalizada é obtido com o acréscimo do termo $\frac{\alpha}{2}\ \mathbf{g}^T\mathbf{Kg}$ em (3.4), ficando assim

$$l_p(\boldsymbol{\theta}) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ng})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ng}) - \frac{\alpha}{2}\mathbf{g}^{\mathbf{T}}\mathbf{Kg}$$

Agora estamos prontos para calcular os estimadores de máxima verossimilhança de $\boldsymbol{\theta}$.

Começando pelo estimador de máxima verossimilhança (EMV) de $\boldsymbol{\beta}$

$\frac{\partial l_p(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = 2\left(\frac{-1}{2\sigma^2}\right)(-\mathbf{X}^T)(\mathbf{Y}\ \text{-}\ \mathbf{X}\boldsymbol{\beta}\ \text{-}\ \mathbf{Ng}) = \frac{1}{\sigma^2}(\mathbf{X}^T)(\mathbf{Y}\ \text{-}\ \mathbf{X}\boldsymbol{\beta}\ \text{-}\ \mathbf{Ng}) = 0,$

$$\mathbf{X^T Y} - \mathbf{X^T X}\hat{\boldsymbol{\beta}} - \mathbf{X}^T\mathbf{N}\hat{\mathbf{g}} = 0,$$
$$\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T(\mathbf{Y} \text{ - } \mathbf{N}\hat{\mathbf{g}}),$$
$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{Y} \text{ - } \mathbf{N}\hat{\mathbf{g}})$$

e

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{Y} \text{ - } \mathbf{N}\hat{\mathbf{g}})$$

Note que, no EMV de $\boldsymbol{\beta}$ a parte $(\mathbf{Y} \text{ - } \mathbf{N}\hat{\mathbf{g}})$ é o resíduo não-paramétrico que é explicado por $\hat{\boldsymbol{\beta}}$.

Agora para $\mathbf{g}$, temos

$$\frac{\partial l_p(\theta)}{\partial \mathbf{g}} = 2\left(\tfrac{-1}{2\sigma^2}\right)(-\mathbf{N}^T)(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}}) - \alpha\hat{\mathbf{g}}\mathbf{K} = 0,$$
$$\mathbf{N}^T(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}}) - \sigma^2\alpha\hat{\mathbf{g}}\mathbf{K} = 0,$$
$$\mathbf{N}^T\mathbf{Y} \text{ - } \mathbf{N}^T\mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}^T\mathbf{N}\hat{\mathbf{g}} - \sigma^2\alpha\hat{\mathbf{g}}\mathbf{K} = 0,$$
$$(\mathbf{N}^T\mathbf{N} + \sigma^2\alpha\mathbf{K})\hat{\mathbf{g}} = \mathbf{N}^T\mathbf{Y} - \mathbf{N}^T\mathbf{X}\hat{\boldsymbol{\beta}}$$

e

$$\hat{\mathbf{g}} = (\mathbf{N}^T\mathbf{N} + \sigma^2\alpha\mathbf{K})^{-1}\mathbf{N}^T(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}})$$

O mesmo ocorre aqui, a parte $(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}})$ do EMV de $\mathbf{g}$ é o resíduo paramétrico que é explicado por $\hat{\mathbf{g}}$.

E por último, vamos calcular o EMV para σ^2

$$\frac{\partial l_p(\boldsymbol{\theta})}{\partial \sigma^2} = -\tfrac{n}{2\hat{\sigma^2}} + \tfrac{1}{2\hat{\sigma^4}}(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}})^T(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}}) = 0$$
$$\hat{\sigma^2} = \tfrac{(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}})^T\ (\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}})}{n}$$

Obtemos os mesmos resultados para $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{g}}$ ao trabalharmos com mínimos quadrados penalizados.

3.4 MATRIZ DE INFORMAÇÃO DE FISHER

A matriz de informação de Fisher é definida da seguinte maneira

$$\text{MI}(\boldsymbol{\theta}) = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}$,$\mathbf{g}$,$\sigma^2)$ e $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Obtemos

$$MI = \begin{bmatrix} I_{\boldsymbol{\beta}\boldsymbol{\beta}} & I_{\boldsymbol{\beta}g} & I_{\boldsymbol{\beta}\sigma^2} \\ I_{g\boldsymbol{\beta}} & I_{gg} & I_{g\sigma^2} \\ I_{\sigma^2\boldsymbol{\beta}} & I_{\sigma^2 g} & I_{\sigma^2\sigma^2} \end{bmatrix}$$

$I_{\boldsymbol{\beta}\boldsymbol{\beta}} = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} = \frac{1}{\hat{\sigma^2}}\mathbf{X}^T\mathbf{X}$,

$I_{\boldsymbol{\beta}_g} = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \mathbf{g}} = \frac{1}{\hat{\sigma^2}}\mathbf{N}^T\mathbf{X}$,

$I_{\boldsymbol{\beta}\sigma^2} = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma^2} = \frac{1}{(\hat{\sigma^2})^2}\mathbf{X}^T(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}})$,

$I_{g\boldsymbol{\beta}} = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \mathbf{g} \partial \boldsymbol{\beta}} = \frac{1}{\hat{\sigma^2}}\mathbf{X}^T\mathbf{N}$,

$I_{gg} = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \mathbf{g}^T \partial \mathbf{g}} = \frac{1}{\hat{\sigma^2}}\mathbf{N}^T\mathbf{N} + \alpha\mathbf{K}$,

$I_{g\sigma^2} = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \mathbf{g} \partial \sigma^2} = \frac{1}{(\hat{\sigma^2})^2}\mathbf{N}^T(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}})$,

$I_{\sigma^2\boldsymbol{\beta}} = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \sigma^2 \partial \boldsymbol{\beta}} = \frac{1}{(\hat{\sigma^2})^2}(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}})^T\mathbf{X}$,

$I_{\sigma^2 g} = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \sigma^2 \partial g} = \frac{1}{(\hat{\sigma^2})^2}(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}})^T\mathbf{N}$

e

$I_{\sigma^2\sigma^2} = -\frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \sigma^2 \partial \sigma^2} = -\frac{n}{2(\hat{\sigma^2})^2} + \frac{1}{(\hat{\sigma^2})^3}(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}})^T(\mathbf{Y} \text{ - } \mathbf{X}\hat{\boldsymbol{\beta}} \text{ - } \mathbf{N}\hat{\mathbf{g}})$

Através da matriz esperada de informação de Fisher obtemos os desvios-padrão dos estimadores, fazendo $\mathrm{DP}(\hat{\boldsymbol{\theta}}) = diag([MI^{-1}]^{1/2})$.

3.5 VALIDAÇÃO CRUZADA PARA MODELO SEMIPARAMÉTRICO

O escore da validação cruzada é dado por

$$CV(\alpha) = \sum_{i=1}^{n}\left(\frac{\mathbf{Y_i} - \hat{\mathbf{Y}}_\mathbf{i}}{1 - \mathbf{A_{ii}}}\right)^2$$

em que $\hat{\mathbf{Y}} = \mathbf{AY} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{N}\hat{\mathbf{g}}$.

A matriz $\mathbf{A}$ é dada por

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} & \mathbf{N} \end{bmatrix} \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{N} \\ \mathbf{N}^T\mathbf{X} & \mathbf{N}^T\mathbf{X} + \sigma^2\alpha\mathbf{K} \end{bmatrix} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{N}^T \end{bmatrix} \tag{3.5}$$

O escore da validação cruzada generalizada possui a forma

$$\mathrm{GCV}(\alpha) = \frac{\sum(\mathbf{Y}_i \text{ - } \hat{Y}_i)^2}{(1 - n^{-1}tr(\mathbf{A}))^2}$$

em que $\mathbf{A}$ é definida da mesma forma como foi definida na validação cruzada e

$\mathbf{N}\hat{\mathbf{g}} = \mathbf{S}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$

$\mathbf{S} = \mathbf{N}(\mathbf{N}^T\mathbf{N} + \alpha\mathbf{K})^{-1}\mathbf{N}^T$

$\mathrm{tr}(\mathbf{A}) = tr(\mathbf{S}) + tr[\{\mathbf{X}^T(\mathbf{I} \text{ - } \mathbf{S})\mathbf{X}\}^{-1}\mathbf{X}^T(\mathbf{I} \text{ - } \mathbf{S})^2\mathbf{X}]$

3.6 MÉTODOS COMPUTACIONAIS PARA ESTIMAÇÃO DOS PARÂMETROS σ^2, $\boldsymbol{\beta}$ E DA CURVA $\mathbf{g}$

Trabalhamos com dois métodos para a estimação do vetor $\boldsymbol{\beta}$, o escalar σ^2 e da curva $\mathbf{g}$. Estes dois métodos são: o método de Green e Silverman[7] (1994) e o método de Eilers e Marx[4] (1996). Eles basicamente se diferenciam na forma de escolha dos nós que serão usados para a estimação da curva $\mathbf{g}$. Falaremos mais detalhadamente nas subseções a seguir.

3.6.1 MÉTODO DE GREEN E SILVERMAN

O método de Green e Silverman já foi introduzido no capítulo "Modelo Não Paramétrico", pois todo capítulo foi baseado na obra de Green e Silverman[7] (1994). Então, nesta seção iremos resumir de forma rápida este método.

Considere $t_1, t_2, ..., t_n$ todos os pontos pertencentes ao vetor $\mathbf{t}$. O método de Green and Silverman selecionará os pontos que ajudarão na estimação da curva $\mathbf{g}$ como sendo os pontos distintos e ordenados pertencentes ao intervalo $t_1, t_2, ..., t_n$. Estes pontos ordenados e distintos serão os nós, que denotaremos por $s_1, ..., s_q$.

Como vimos na seção 3.2, construímos a matriz de incidência $\mathbf{N}$ (dimensão nxq) com entradas $N_{ij} = 1$ se $t_i = s_j$ e 0 caso contrário. Obtemos também, de acordo com a seção 2.2 as matrizes $\mathbf{Q}$, $\mathbf{R}$ e $\mathbf{K}$, lembrando que agora os t_i serão os s_i.

Iremos expor agora um passo a passo do código utilizado para a estimação do vetor $\boldsymbol{\beta}$, do escalar σ^2 e da curva $\mathbf{g}$ através do método de Green and Silverman[7] (1994).

Passo 1:

- O primeiro passo é a construção da matriz de incidência $\mathbf{N}$, através dos elementos t_1, t_2, ..., t_n e $s_1, ..., s_q$.
- Depois, construímos as matrizes $\mathbf{Q}$, $\mathbf{R}$ e $\mathbf{K}$, com base nos $s_1, ..., s_q$, que nos possibilita construirmos a matriz $\mathbf{K}$.

Passo 2:

- Criamos uma determinada função onde entramos com $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{N}$, $\mathbf{K}$ e um intervalo para α. Dentro desta função chamamos outra função que irá calcular o valor ótimo de α, ou seja, esta função que chamamos é a da validação cruzada.

Passo 3:

- Como até este passo temos o valor ótimo de α e as matrizes $\mathbf{Q}$, $\mathbf{R}$ e $\mathbf{K}$, podemos por fim calcular as estimativas para $\boldsymbol{\beta}$, σ^2 e $\mathbf{g}$.

- Nesta função entramos com **Y, X, N, K** e α.
- E dentro desta função, consideramos os seguintes valores iniciais para $\boldsymbol{\beta}$, σ^2 e **g**:

  $\boldsymbol{\beta}_0 = \mathbf{(X^TX)^{-1}XY}$

  $\sigma_0^2 = \mathbf{(Y - X\boldsymbol{\beta}_0)^T(Y - X\boldsymbol{\beta}_0)}/n$

  $\mathbf{g} = (\mathbf{N^TN} + \alpha\ \sigma_0^2\mathbf{K})\mathbf{N^T(Y - X\boldsymbol{\beta}_0)}$

  Guardamos estes valores em $\boldsymbol{\theta}_0$.

  Criamos:

  $criterio = 1$

  $cont = 0$ e utilizamos a função while com os seguintes comandos de parada:

  $criterio < 10^{-5}$ e $cont < 5000$.

  E dentro deste while vamos atualizar os valores de $\boldsymbol{\beta}$, σ^2 e **g** para

  $\boldsymbol{\beta} = \mathbf{(X^TX)^{-1}X(Y - Ng)}$

  $\sigma^2 = \mathbf{(Y - X\boldsymbol{\beta} - Ng)^T(Y - X\boldsymbol{\beta} - Ng)}/n$

  $\mathbf{g} = (\mathbf{N^TN} + \alpha\ \sigma^2\mathbf{K})\mathbf{N^T(Y - X\boldsymbol{\beta})}$

  Guardamos os valores atualizados em $\boldsymbol{\theta}$

  criterio = $||\boldsymbol{\theta}_0 - \boldsymbol{\theta}||$

  Essa função retorna $\boldsymbol{\theta}$.
- Assim, temos as estimativas para os parâmetros $\boldsymbol{\beta}$, σ^2 e a curva **g**.

3.6.2 MÉTODO DE EILERS E MARX

Basicamente o que muda de um método para o outro é a escolha dos nós.

Nesta seção, iremos ver que no método de Eilers e Marx[4] (1996) os nós são escolhidos de acordo com um intervalo pré estabelecido usando a ideia dos B-splines. B-splines são construídos a partir de peças polinomiais, unidas a certos valores dos nós. Segue algumas propriedades dos B-splines de grau q

-Consiste em $q + 1$ polinômios de grau q,

-Pode-se calcular os B-splines,

-As peças polinomiais se unem em q nós internos,

-Dividir o domínio t_{min} a t_{max} em n' intervalos iguais para n'+1 nós,

-Cada intervalo será coberto por $q + 1$ B-splines de grau q

-Total de nós será $n' + 2q + 1$,

-O número de B-splines será $n = n' + q$,

-Os nós serão equidistantes e

-Uma combinação linear de B-splines de terceiro grau dá uma curva suave.

A matriz incidente de Eilers e Marx[4] (1996) é denotada por **B**.

Seja B_j(t,q), denote o valor em t do j-ésimo B-spline com grau q para um certo número de nós equidistantes. A curva ajustada $\hat{\mathbf{g}}$ para dados (t_i, y_i) é a combinação linear $\hat{g}(t) = \sum_{j=1}^{n} \hat{a_j} B_j(t, q)$.

Considere a regressão de m pontos de dados (t_i, y_i) em um conjunto de n B-splines $B_j(.)$. A função dos mínimos quadrados é a seguinte

$$S = \sum_{i=1}^{m} \{y_i - \sum_{j=1}^{n} a_j B_j(t_i)\}^2$$

Para tornar a função acima mais flexível O'Sullivan[9] (1986) introduziu uma penalidade na segunda derivada da curva ajustada e assim formando a função

$$S = \sum_{i=1}^{m} \{y_i - \sum_{j=1}^{n} a_j B_j(t_i)\}^2 + \alpha \int_{t_{min}}^{t_{max}} \{\sum_{j=1}^{n} a_j B_j^{(2)}(t)\}^2 dt \tag{3.6}$$

De Boor[2] (1978) dá uma maneira simples de calcular a derivadas de B-splines

$$h^2 \sum_j a_j B_j^{(2)}(x; q) = \sum_j \Delta^2 a_j B_j(x; q-2)$$

em que h é a distância entre os nós e $\Delta a_j = a_j - a_{j-1}$. Assim, podemos reescrever (3.6)

$$S = \sum_{i=1}^{m} \{y_i - \sum_{j=1}^{n} a_j B_j(t_i)\}^2 + \alpha \sum_{j=k+1}^{n} (\Delta^k a_j)^2 \tag{3.7}$$

O sistema de equação que se segue da minimização de (3.7) pode ser escrito como

$$\mathbf{B^T y} = (\mathbf{B^T B} + \alpha \mathbf{D_k^T D_k})\hat{\mathbf{a}}$$
$$\hat{\mathbf{a}} = (\mathbf{B^T B} + \alpha \mathbf{D_k^T D_k})^{-1} \mathbf{B^T y}$$

em que $\mathbf{D}_k$ é a representação da matriz do operador da diferença Δ_k e os elementos de **B** são $b_{ij} = B_j(x_i)$. Fazendo uma analogia ao método de Green e Silverman, $\hat{\mathbf{a}}$ seria a curva $\hat{g}$ estimada. E $\mathbf{D}^T\mathbf{D} = \mathbf{M}$, que seria análoga a matriz **K** de Green e Silverman.

Considere o modelo semiparamétrico $\mathbf{Y = X\beta + Ba + erro}$, com n observações independentes e parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{a}, \sigma^2)$, em que **Y** segue uma distribuição normal

com valor esperado igual a $\mathbf{X}\boldsymbol{\beta} + \mathbf{Ba}$ e variância σ^2. Podemos escrever a máxima verossimilhança penalizada na forma

$$l_p(\boldsymbol{\theta}) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ba})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Ba}) - \frac{\alpha}{2}\mathbf{a^T Ma} \quad (3.8)$$

Nota-se que a única diferença entre os dois métodos na função de máxima verossimilhança penalizada é que aqui $\mathbf{N}$, $\mathbf{g}$ e $\mathbf{K}$ equivalem a $\mathbf{B}$, $\mathbf{a}$ e $\mathbf{M}$. As expressões para os cálculos dos estimadores de máxima verossimilhança e os resultados de (3.8) são análogos aos da seção 3.3.

O escore da validação cruzada é dado por

$$CV(\alpha) = \sum_{i=1}^{m}\left(\frac{Y_i - \hat{Y}_i}{1 - h_{ii}}\right)^2$$

em que $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{B}\hat{\mathbf{a}} = \mathbf{HY}, \quad \mathbf{H} = \mathbf{B}(\mathbf{B}^\top\mathbf{B} + \alpha\mathbf{D}_k^\top\mathbf{D}_k)^{-1}\mathbf{B}^\top$.

Segue o score da validação cruzada generalizada

$$GCV(\alpha) = \sum_{i=1}^{m}\frac{(Y_i - \hat{Y}_i)^2}{(m - \sum_{i=1}^{m} h_{ii})^2}$$

Depois desta introdução ao método de Eilers e Marx (1996), vamos iniciar a explicação passo a passo do algoritmo, mas lembrando que há somente uma diferença entre os códigos, que é na construção da matriz $\mathbf{B}$ e consequentemente na matriz $\mathbf{M}$, o restante do código segue semelhante ao método de Green e Silverman (1994).

Passo 1:

- Para a construção da matriz $\mathbf{B}$ precisamos entrar com os seguintes argumentos: o argumento da variável não-paramétrica, $\mathbf{t}$, ndx=número de intervalos, bdeg=grau do B-spline (em nossos estudos trabalhamos com bdeg=3), min($\mathbf{t}$) e max($\mathbf{t}$).

- Nisso, dentro da função temos que calcular o comprimento de cada intervalo, mas como serão intervalos equidistantes, eles terão o mesmo comprimento. O comprimento do intervalo será denotado por dx e será calculado, por: $dx = (max(\mathbf{t}) - min(\mathbf{t}))/ndx$.

- Agora, calculamos os nós da seguinte forma

  $knots = seq(xl - bdeg * dx, xr + bdeg * dx, by = dx)$

  Eilers e Marxs calculam também bdeg nós externos, para suavizar o final de cada curva estimada. Por isso esse termo xl-bdeg*dx e xr+bdeg*dx, ou seja, ele está

calculando os nós externos na parte esquerda e direita da curva, fora do intervalo t. Estes nós externos possuem o mesmo comprimento dos nós internos. O comando seq nos dá uma sequência de números neste intervalo $xl - bdeg * dx, xr + bdeg * dx$ (inclusive) sendo que a distância de um para outro é igual a dx.

- Por fim, usa-se o comando

  $B = splineDesign(knots, x, bdeg + 1, 0 * x, outer.ok = T)$

  que nos retornará a matriz **B**. Como se vê, entramos com os nós, o vetor **t** e o $grau + 1$. Cada linha da matriz **B** será formada pelos coeficientes do polinômio referente a cada t_i.

  Passo 2:

- Agora, construímos a matriz **D** dá seguinte forma

  D<-diag(ncol(B)) for (k in 1:2) D<-diff(D)

  E obtemos a matriz **M**, $\mathbf{M} = \mathbf{D}^T\mathbf{D}$

- O restante do algoritmo segue o mesmo procedimento a partir do Passo 2 de Green e Silverman.

3.7 SIMULAÇÃO

Nesta seção, iremos apresentar simulações feitas para os modelos semiparamétricos normais utilizando os métodos de Green e Silverman[7] (1994) e Eilers e Marx[4] (1996). O software utilizado foi o R Core Team[11] (2017).

Primeiro, considere o modelo

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Ng} + \epsilon$$

em que, $\epsilon \sim N(0, \sigma^2\mathbf{I})$ e $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Ng}, \sigma^2\mathbf{I})$.

Temos:

-Três variáveis explicativas paramétricas: $x_1 \sim N(0; 0, 25)$, $x_2 \sim N(0, 1)$ e $x_3 \sim U(0, 2)$. Todas de tamanho igual a 100 ($n = 100$). Assim, temos uma matriz **X** de dimensão 100x3.

-t_1: uma sequência de 50 valores no intervalo de (0,4π)

-t: vetor de tamanho 100 onde os elementos são os valores do vetor t_1 repetidos uma vez, ou seja, uma junção de t_1 com t_1. Estes valores estão ordenados.

- $g(t) = sin(t)$

- $\boldsymbol{\epsilon} \sim N(0; 0.25)$

-$\boldsymbol{\beta}_1 = 1$, $\boldsymbol{\beta}_2 = -1$ e $\boldsymbol{\beta}_3 = 1$

-$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g}(\mathbf{t}) + \boldsymbol{\epsilon}$

-Intervalo para α: 0.1 a 4

Obtemos as estimativas pelo método de Green and Silverman na Tabela 1.

Tabela 1 – Estimativas encontradas por Green and Silverman para modelo semiparamétrico

	Estimativa	Erro Padrão
$\boldsymbol{\beta}_1$	0,9282	0,1053
$\boldsymbol{\beta}_2$	-0,9511	0,0481
$\boldsymbol{\beta}_3$	1,0794	0,0989
σ^2	0,2507	0,0361
α	2,5146	-

E a curva $\hat{\mathbf{g}}$ estimada, em cinza contínuo e as bandas de confiança delimitando a curva na Figura 3.

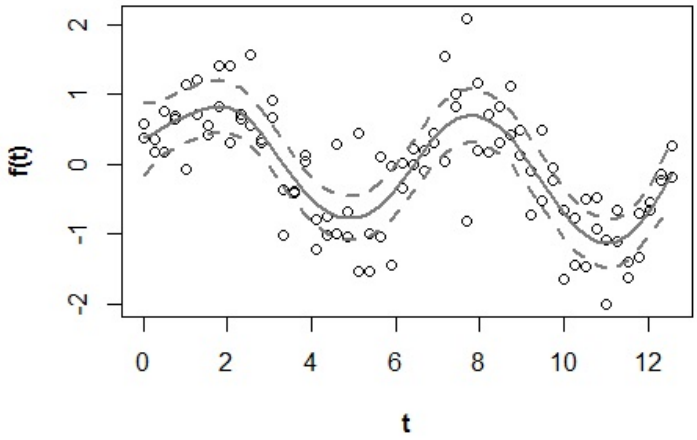


Figura 3 – Curva estimada $\hat{\mathbf{g}}$ através do método de Green and Silverman para modelo semiparamétrico

Pelo método de Eilers e Marx temos apenas que entrar a mais na simulação com a quantidade de intervalos, que neste caso entramos com 14, e o grau da função B-spline, que entramos com grau igual a 3.

Obtemos as estimativas na Tabela 2.

E a curva $\hat{\mathbf{g}}$ estimada, em cinza contínuo, cercada pela banda de confiança na Figura 4.

Tabela 2 – Estimativas encontradas por Eilers and Marx para modelo semiparamétrico

	Estimativa	Erro Padrão
$\boldsymbol{\beta}_1$	0,9241	0,1243
$\boldsymbol{\beta}_2$	-1,0149	0,0623
$\boldsymbol{\beta}_3$	1,0791	0,1001
σ^2	0,3185	0,0457
α	1,9551	-

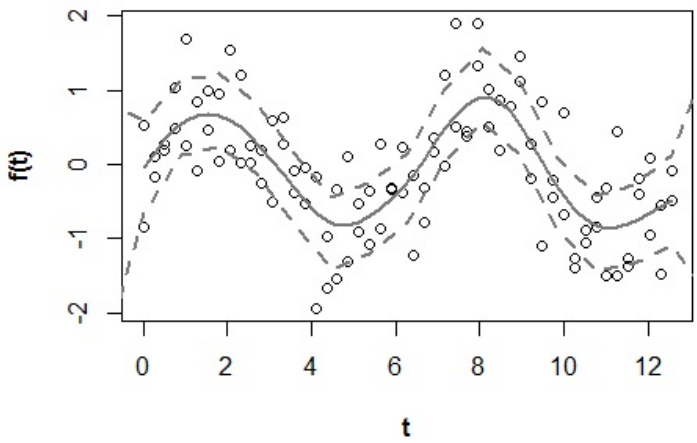


Figura 4 – Curva estimada $\hat{\mathbf{g}}$ através do método de Eilers and Marx para modelo semiparamétrico

Observamos que não há diferenças significativas nas estimativas para os $\boldsymbol{\beta}$ de um método para outro, os erros padrão são próximos também. A variável não paramétrica está bem ajustada para os dois métodos, acompanhando os resíduos não paramétricos do modelo. Somente no valor da estimativa de α que se tem uma diferença de aproximadamente uma unidade de um método para o outro. Green e Silverman[7] (1994) apresentando um valor maior para está estimativa. Isto pode estar relacionado a maneira de escolha dos nós, pois Eilers e Marx[4] (1996) acabam colhendo mais características, informações sobre a curva da maneira que escolhem os nós, com isso, precisam de um valor de α menor para suavizá-la.

4 MODELO BINOMIAL SEMIPARAMÉTRICO

Ao trabalharmos com Modelos Lineares Generalizados abrimos nosso leque de distribuições ao qual a variável resposta Y possa pertencer. Agora, ela pode seguir todas as distribuições pertencentes à família exponencial.

No modelo binomial, a variável resposta Y representa o número de sucessos em m tentativas(ensaios) de Bernoulli(μ), mas em nossos estudos optamos por trabalhar com a variável resposta Y* sendo a proporção de sucessos em m ensaios independentes, cada um com probabilidade de ocorrência igual a μ. Utilizamos o número de tentativas fixo, mas isto poderia variar, assim teríamos um vetor de diferentes números de ensaios. Para contextualizar, a seguir alguns exemplos para a variável Y: número de sementes que germinam depois de m serem plantadas, quantidade de peças não conformes em m ensaios, número de cupons usados entre os m distribuídos.

Assumimos que mY* $\sim$ Binomial(m,μ). Seja $Y_1^*, Y_2^*, ..., Y_n^*$ uma amostra aleatória de uma binomial, a função de probabilidades de Y* fica assim definida

$$P(mY^* = my^*) = \binom{m}{my^*}\mu^{my^*}(1-\mu)^{m-my^*} = \binom{m}{my^*}\left(\frac{\mu}{1-\mu}\right)^{my^*}(1-\mu)^m \qquad (4.1)$$

Utilizando transformações em (4.1), temos

$$P(mY^* = my^*) = exp\left\{log\binom{m}{my^*} + my^* log\left(\frac{\mu}{1-\mu}\right) + mlog(1-\mu)\right\}$$

$$P(mY^* = my^*) = exp\left\{m\left[y^* log\left(\frac{\mu}{1-\mu}\right) + log(1-\mu)\right] + log\binom{m}{my^*}\right\} \qquad (4.2)$$

A função em (4.2) está na forma das distribuições da família exponencial, dada por

$$f(y;\boldsymbol{\theta},\phi) = exp\left[\phi\left\{y\boldsymbol{\theta} - b(\boldsymbol{\theta})\right\} + c(y,\phi)\right], \qquad (4.3)$$

em que $\phi = m$, $\boldsymbol{\theta} = log\left(\frac{\mu}{1-\mu}\right)$, $b(\boldsymbol{\theta}) = -log(1-\mu)$ e c$(y^*, \phi) = log\binom{m}{my^*}$.

$\boldsymbol{\theta} = log\left(\frac{\mu}{1-\mu}\right) \Leftrightarrow \mu = \frac{e^{\boldsymbol{\theta}}}{1+e^{\boldsymbol{\theta}}}$,

$b(\boldsymbol{\theta}) = -log(1-\mu) = -log\left(\frac{1}{1+e^{\boldsymbol{\theta}}}\right) = log(1+e^{\boldsymbol{\theta}})$

e

$V(\mu) = \frac{d\mu_i}{d\boldsymbol{\theta}_i} = \frac{e^{\boldsymbol{\theta}}}{(1+e^{\boldsymbol{\theta}})^2} \Leftrightarrow V(\mu) = \mu(1-\mu)$

Os modelos lineares generalizados são definidos por (4.3) e pela parte sistemática

$$g(\mu_i) = \eta_i.$$

Para o caso da regressão logística semiparamétrica o MLG é definido assim

$$log\left(\frac{\mu_i}{1-\mu_i}\right) = \mathbf{x}_i^T\boldsymbol{\beta} + g(t_i) \tag{4.4}$$

em que $g(\mu_i)$ é uma função monótona e diferenciável, denominada função de ligação *logit*. Ela é uma função de ligação canônica, mas existem outros tipos de ligações que podem ser usadas e que falaremos mais adiante neste trabalho. $\mathbf{x_i^T}$ vetor de covariáveis paramétricas e $\mathbf{n_i^T}$ vetor da matriz incidente. Assim, $g(\mu_i) = log\left(\frac{\mu_i}{1-\mu_i}\right)$ e $\eta_i = \mathbf{x_i^T}\boldsymbol{\beta} + \mathbf{n_i^T}\mathbf{g}$.

Quando trabalhamos com o valor esperado de Y*, temos que E(Y*) = μ, em que μ é definido no intervalo [0,1], mas como a parte direita da equação (4.4) está definida para toda a reta, temos que utilizar alguma transformação neste valor esperado para que ele também passe a ser definido em toda a reta, pois assim não haverá nenhuma restrição para a estimação dos parâmetros $\boldsymbol{\beta}$ e da curva g. Então, utilizamos o $\log(\mu/(1-\mu))$. Esta função que aplicamos em μ é a função de ligação, como dito antes a função de ligação *logit* não é a única, podemos aplicar outros tipos de funções em μ, como a função de ligação *probit* e complemento log-log ou outra qualquer que faça μ ser definido em toda a reta.

4.1 MÁXIMA VEROSSIMILHANÇA PENALIZADA PARA O MODELO BINOMIAL SEMIPARAMÉTRICO

Considere

$$f(y_i;\boldsymbol{\theta}_i,\phi) = exp\left[\phi\left\{y_i\boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)\right\} + c(y_i,\phi)\right].$$

O logaritmo de máxima verossimilhança penalizada para o modelo logístico semiparamétrico é definido assim

$$l(\boldsymbol{\beta},\mathbf{g}) = m\sum_{i=1}^{n}\left[y_i^* log\left(\frac{\mu_i}{1-\mu_i}\right) + log(1-\mu_i)\right] + \sum_{i=1}^{n} log\binom{m}{my_i^*} - \frac{\alpha}{2}\mathbf{g^T K g}, \tag{4.5}$$

em que acrescentamos o termo (a *penalty*) que incorpora as restrições para a estimação da curva $\mathbf{g}$, a mesma ideia utilizada para o modelo normal semiparamétrico.

4.1.1 FUNÇÃO ESCORE E INFORMAÇÃO DE FISHER

Iremos considerar $\boldsymbol{\theta} = (\boldsymbol{\beta},\mathbf{g})$ e $l(\boldsymbol{\theta})$ da definição de máxima verossimilhança penalizada em (4.5).

Para obtermos a função escore, temos primeiro que calcular as derivadas parciais de $l(\boldsymbol{\theta})$ em relação a $\boldsymbol{\beta}$ e a $\mathbf{g}$.

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}_j} = m \sum_{i=1}^{n} \left\{ y_i^* \frac{d\boldsymbol{\theta}_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}_j} - \frac{db(\boldsymbol{\theta}_i)}{d\boldsymbol{\theta}_i} \frac{d\boldsymbol{\theta}_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}_j} \right\},$$

onde

$\frac{d\boldsymbol{\theta}_i}{d\mu_i} = \frac{1}{\mu_i(1-\mu_i)} = [V(\mu_i)]^{-1}$,

$\dfrac{\partial \eta_i}{\partial \boldsymbol{\beta}_j} = \mathbf{x_i}$ e

$\frac{db(\boldsymbol{\theta}_i)}{d\boldsymbol{\theta}_i} = \frac{e^{\boldsymbol{\theta}_i}}{1+e^{\boldsymbol{\theta}_i}} = \mu_i$, com

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}_j} = m \sum_{i=1}^{n} \left\{ y_i^* \left[V(\mu_i)\right]^{-1} \frac{d\mu_i}{d\eta_i} \mathbf{x_i} - \mu_i \left[V(\mu_i)\right]^{-1} \frac{d\mu_i}{d\eta_i} \mathbf{x_i} \right\}$$

e

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}_j} = m \sum_{i=1}^{n} \left\{ (y_i^* - \mu_i) \left[V(\mu_i)\right]^{-1} \frac{d\mu_i}{d\eta_i} \mathbf{x_i} \right\}.$$

Considerando $w_i = \frac{(d\mu_i/d\eta_i)^2}{V_i}$, temos

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}_j} = m \sum_{i=1}^{n} \left\{ \sqrt{\frac{w_i}{V_i}} (y_i^* - \mu_i) \mathbf{x_i} \right\}.$$

Podemos escrever a função escore em relação a $\boldsymbol{\beta}$, na forma matricial, dada abaixo

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}} = m \mathbf{X^T W^{\frac{1}{2}} V^{-\frac{1}{2}} (Y^* - \boldsymbol{\mu})}$$

Como estamos trabalhando com a ligação canônica, temos que $\boldsymbol{\theta} = \eta$, ou seja, $log\,(\mu/1-\mu) = \eta$, assim

$w_i = \frac{(d\mu_i/d\eta_i)^2}{V_i}$ e

$\frac{d\mu_i}{d\eta_i} = \left[g^{'}(\mu_i)\right]^{-1} = \mu_i(1-\mu_i)$. Note que

$V_i = \mu_i(1-\mu_i)$ e então,

$w_i = \frac{\mu_i^2(1-\mu_i)^2}{\mu_i(1-\mu_i)} = \mu_i(1-\mu_i)$.

Assim temos que $\mathbf{W}$, que é denotada matriz de pesos, é igual a matriz $\mathbf{V}$.

$\mathbf{W} = diag(w_1, ..., w_n)$ e $\mathbf{V} = diag(V_1, ..., V_n)$

Portanto, temos que

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}} = m\mathbf{X}^{\mathbf{T}}(\mathbf{Y}^* - \boldsymbol{\mu}).$$

Agora em relação a $\mathbf{g}$, temos que

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g}} = m\sum_{i=1}^{n}\left\{y_i^*\frac{d\boldsymbol{\theta}_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial g} - \frac{db(\boldsymbol{\theta}_i)}{d\boldsymbol{\theta}_i}\frac{d\boldsymbol{\theta}_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial g}\right\} - \frac{d(\mathbf{g}^{\mathbf{T}}\mathbf{K}\mathbf{g})}{d\mathbf{g}}.$$

Note que a única diferença será na derivada $\dfrac{\partial \eta_i}{\partial \mathbf{g}} = \mathbf{n_i}$, então

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g}} = m\sum_{i=1}^{n}\left\{y_i^*\left[V(\mu_i)\right]^{-1}\frac{d\mu_i}{d\eta_i}\mathbf{n_i} - \mu_i\left[V(\mu_i)\right]^{-1}\frac{d\mu_i}{d\eta_i}\mathbf{n_i}\right\} - \alpha\mathbf{K}\mathbf{g},$$

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g}} = m\sum_{i=1}^{n}\left\{(y_i^* - \mu_i)\left[V(\mu_i)\right]^{-1}\frac{d\mu_i}{d\eta_i}\mathbf{n_i}\right\} - \alpha\mathbf{K}\mathbf{g}$$

e

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g}} = m\sum_{i=1}^{n}\left\{\sqrt{\frac{w_i}{V_i}}(y_i^* - \mu_i)\mathbf{n_i}\right\} - \alpha\mathbf{K}\mathbf{g}.$$

Na forma matricial, temos que

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g}} = m\mathbf{N}^{\mathbf{T}}\mathbf{W}^{\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}}(\mathbf{Y}^* - \boldsymbol{\mu}) - \alpha\mathbf{K}\mathbf{g}$$

e

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g}} = m\mathbf{N}^{\mathbf{T}}(\mathbf{Y}^* - \boldsymbol{\mu}) - \alpha\mathbf{K}\mathbf{g}.$$

Concluindo a obtenção da função escore que denotamos por

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}} \\ \dfrac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g}} \end{bmatrix},$$

tal que

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} m\mathbf{X}^{\mathbf{T}}(\mathbf{Y}^* - \boldsymbol{\mu}) \\ m\mathbf{N}^{\mathbf{T}}(\mathbf{Y}^* - \boldsymbol{\mu}) - \alpha\mathbf{K}\mathbf{g} \end{bmatrix} \tag{4.6}$$

Para obter a matriz de informação de Fisher, precisamos das derivadas segundas de $l(\boldsymbol{\beta}, \mathbf{g})$ em relação a $\boldsymbol{\beta}$ e a $\mathbf{g}$. Temos,

$$log\left(\frac{\mu_i}{1-\mu_i}\right) = \mathbf{x_i^T}\boldsymbol{\beta} + \mathbf{n_i^T}\mathbf{g},$$

onde

$$\mu_i = \frac{e^{\mathbf{x_i^T}\boldsymbol{\beta}+\mathbf{n_i^T}\mathbf{g}}}{1+e^{\mathbf{x_i^T}\boldsymbol{\beta}+\mathbf{n_i^T}\mathbf{g}}}.$$

Assim, como já mostramos

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}} = m\mathbf{X^T}(\mathbf{Y}^* - \boldsymbol{\mu})$$

e consequentemente

$$\frac{\partial^2 l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} = -m\mathbf{X^T}\frac{\partial d\boldsymbol{\mu}}{\partial \boldsymbol{\beta}^T},$$

onde

$\frac{\partial d\boldsymbol{\mu}}{\partial \boldsymbol{\beta}^T} = \mathbf{X}\frac{e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}}}{(1+e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}})^2}$. Portanto,

$$\frac{\partial^2 l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} = -m\mathbf{X^T}\frac{e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}}}{(1+e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}})^2}\mathbf{X}.$$

Denotando a matriz peso $\mathbf{W}$ por diagonal de $\frac{e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}}}{(1+e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}})^2}$, temos que

$$\frac{\partial^2 l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} = -m\mathbf{X^T}\mathbf{W}\mathbf{X}.$$

Procedimento análogo, utilizamos para $\mathbf{g}$, tal que

$$\frac{\partial^2 l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g}\partial \mathbf{g^T}} = -m\mathbf{N^T}\frac{\partial d\boldsymbol{\mu}}{\partial \mathbf{g^T}} - \alpha\mathbf{K},$$

onde

$\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{g}^T} = \mathbf{N}\frac{e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}}}{(1+e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}})^2}$. Portanto,

$$\frac{\partial^2 l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g}\partial \mathbf{g^T}} = -m\mathbf{N^T}\frac{e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}}}{(1+e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{Ng}})^2}\mathbf{N} - \alpha\mathbf{K}$$

e então,

$$\frac{\partial^2 l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g} \partial \mathbf{g}^{\mathbf{T}}} = -m\mathbf{N}^{\mathbf{T}}\mathbf{W}\mathbf{N} - \alpha\mathbf{K}.$$

A segunda derivada de $l(\boldsymbol{\beta}, \mathbf{g})$ em relação a $\boldsymbol{\beta}$ e $\mathbf{g}$ é dada por

$$\frac{\partial^2 l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g} \partial \boldsymbol{\beta}^T} = \frac{\partial}{\partial \mathbf{g}} \left(\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}^T} \right),$$

onde

$\frac{\partial l(\boldsymbol{\beta}, \mathbf{g})}{\partial \boldsymbol{\beta}^T} = m\mathbf{X}(\mathbf{Y}^* - \boldsymbol{\mu})$ e consequentemente

$\frac{\partial m\mathbf{X}(\mathbf{Y}^* - \boldsymbol{\mu})}{\partial \mathbf{g}} = -m\mathbf{X}\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{g}} \rightarrow \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{g}} = \mathbf{N}^{\mathbf{T}} \frac{e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{N}\mathbf{g}}}{(1+e^{\mathbf{X}\boldsymbol{\beta}+\mathbf{N}\mathbf{g}})^2}$. Assim,

$$\frac{\partial^2 l(\boldsymbol{\beta}, \mathbf{g})}{\partial \mathbf{g} \partial \boldsymbol{\beta}^T} = -m\mathbf{N}^{\mathbf{T}}\mathbf{W}\mathbf{X}$$

Podemos definir a matriz de informação esperada de Fisher da seguinte forma

$$\mathbf{K}(\boldsymbol{\theta}) = \begin{bmatrix} m\mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{X} & m\mathbf{X}^{\mathbf{T}}\mathbf{W}\mathbf{N} \\ m\mathbf{N}^{\mathbf{T}}\mathbf{W}\mathbf{X} & m\mathbf{N}^{\mathbf{T}}\mathbf{W}\mathbf{N} + \alpha\mathbf{K} \end{bmatrix} \tag{4.7}$$

A função escore e matriz esperada de informação de Fisher serão utilizadas na estimação dos parâmetros pelo método de Newton-Raphson. Através da matriz esperada de informação de Fisher também obtemos os desvios padrão dos parâmetros.

4.2 ESTIMAÇÃO DE $\boldsymbol{\beta}$ E $\mathbf{g}$

Para os modelos lineares generalizados não é possível, exceto quando se trabalha com a variável resposta seguindo uma distribuição normal, chegar em uma expressão analítica para os estimadores de máxima verossimilhança de $\boldsymbol{\beta}$ e $\mathbf{g}$, como foi feito na seção 3.3. Assim, devemos utilizar algum método iterativo para obtenção das estimativas, o método aqui utilizado foi o método de Newton-Raphson.

Considere $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathbf{T}}, \mathbf{g}^{\mathbf{T}})^T$, U($\boldsymbol{\theta}$) a função escore obtida em (4.6) e $\mathbf{K}(\boldsymbol{\theta})$ a matriz de informação esperada de Fisher em (4.7). Vamos estimar $\boldsymbol{\theta}$ da seguinte maneira

$$\boldsymbol{\theta}^{(\mathbf{m+1})} = \boldsymbol{\theta}^{(\mathbf{m})} + \left\{ \mathbf{K}^{-\mathbf{1}}(\boldsymbol{\theta}) \right\}^{(m)} \mathbf{U}(\boldsymbol{\theta})^{(\mathbf{m})} \tag{4.8}$$

Em todo método iterativo existem dois pontos que devem ser determinados: estimativa inicial (chute inicial) e o critério de parada.

Utilizamos como chute inicial $\boldsymbol{\theta}^{(m)}$ no algoritmo em (4.8):

$\boldsymbol{\beta}_{\mathbf{0}} = 0$, um vetor de zeros

$\mathbf{g_0} = (\mathbf{N^T N})^{-1}\mathbf{N^T G}$, em que $\mathbf{G}$ é $\mathbf{Y}$ aplicado na função de ligação, ou seja, g($\mathbf{Y}$).

Assim o algoritmo em (4.8) irá calcular o próximo passo $\boldsymbol{\theta}^{(m+1)}$ e continuará sucessivamente até o critério de parada. Utilizamos como critério de parada a função $||\boldsymbol{\theta}^{(m)} - \boldsymbol{\theta}^{(m+1)}|| < 10^{-3}$.

Finalizado o algoritmo teremos as estimativas para $\boldsymbol{\beta}$ e $\mathbf{g}$.

4.3 VALIDAÇÃO CRUZADA

Três critérios podem ser utilizados para a validação cruzada. Um primeiro critério é dado por Green e Silverman[7] (1994), subseção 5.4.3.

$$CV(\alpha) = \sum_{i=1}^{n} \frac{d_i}{(1 - A_{ii})^2} \tag{4.9}$$

em que a deviance será denotada como definida por Paula[10] (2004).

Assumimos que $Y_i \sim Binomial(m, \mu_i)$.

Se $0 < y_i < m$, temos que

$$d_i = 2\sum_{i}^{n} \left\{ y_i(\tilde{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_i) + b(\hat{\boldsymbol{\theta}}_i) - b(\tilde{\boldsymbol{\theta}}_i) \right\},$$

em que $\boldsymbol{\theta}_i = log(\mu_i/1 - \mu_i)$ e $b(\boldsymbol{\theta}_i) = -log(1 - \mu_i)$. Assim,

$\hat{\boldsymbol{\theta}}_i = log(\hat{\mu}_i/1 - \hat{\mu}_i)$, $\tilde{\boldsymbol{\theta}}_i = log\left(\frac{y_i/n}{1-y_i/n}\right)$, $b(\hat{\boldsymbol{\theta}}_i) = -log(1-\hat{\mu}_i)$ e $b(\tilde{\boldsymbol{\theta}}_i) = -log(1-y_i/n)$. Portanto,

$$d_i = 2\sum_{i=1}^{k}[y_i log(y_i/m\hat{\mu}_i) + (m - y_i)log\{(1 - y_i/m)/(1 - \hat{\mu}_i)\}].$$

Se $y_i = 0$, temos

$$d_i = -2mlog(1 - \hat{\mu}_i)$$

Se $y_i = m$, temos

$$d_i = -2mlog\hat{\mu}_i$$

A matriz $\mathbf{A}$ foi definida em (3.5) na seção 3.5, é a matriz referente ao modelo normal semiparamétrico.

Um segundo critério é utilizar somente a deviance na validação cruzada, ou seja, a validação cruzada ficaria definida como

$$CV(\alpha) = \sum_{i=1}^{n} d_i$$

Por último, o terceiro critério seria a estimação de α retirando observações da deviance a cada passo, ou seja, considerando

$$d_i^{(-i)} = 2\left\{y_i(\tilde{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_i^{(-i)}) + b(\hat{\boldsymbol{\theta}}_i^{(-i)}) - b(\tilde{\boldsymbol{\theta}}_i)\right\},$$

temos que

$$CV(\alpha) = \sum_{i=1}^{n} d_i^{(-i)}.$$

Testes foram feitos com cada um dos critérios para se trabalhar com a validação cruzada, mas obtemos resultados parecidos nos três critérios, não teve um critério que obtinha melhores resultados em comparação aos outros, então decidimos utilizar o que fosse mais simples, ou seja, que o algoritmo fosse mais simples e mais rápido. Assim, em toda parte de simulações e aplicação a dados reais, que serão mostrados mais adiante, usamos a validação cruzada denotada por 4.9.

4.4 OUTRAS FUNÇÕES DE LIGAÇÃO: *PROBIT* E COMPLEMENTO LOG-LOG

Seja μ a proporção de sucessos de uma distribuição binomial. A ligação *probit* é definida por

$$\Phi^{-1}(\mu) = \eta,$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão.

A única diferença na função escore e na matriz esperada de informação de Fisher para a ligação *probit* será nas matrizes, $\mathbf{W}$ e $\mathbf{V}$, pois

$$\mu_i = \Phi(\eta_i),$$

então

$$\frac{d\mu_i}{d\eta_i} = \phi(\eta_i),$$

$$V_i = \mu_i(1-\mu_i) = \Phi(\eta_i)[1-\Phi(\eta_i)]$$

e

$$w_i = \frac{(d\mu_i/d\eta_i)^2}{V_i} = \frac{[\phi(\eta_i)]^2}{\Phi(\eta_i)[1-\Phi(\eta_i)]}.$$

A função escore e a matriz esperada de informação de Fisher ficam definidas como

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} m\mathbf{X^T W^{1/2} V^{-1/2}}(\mathbf{Y}^* - \mu) \\ m\mathbf{N^T W^{1/2} V^{-1/2}}(\mathbf{Y}^* - \mu) - \alpha\mathbf{Kg} \end{bmatrix}$$

e

$$K(\boldsymbol{\theta}) = \begin{bmatrix} m\mathbf{X^T WX} & m\mathbf{X^T WN} \\ m\mathbf{N^T WX} & m\mathbf{N^T WN} + \alpha\mathbf{K} \end{bmatrix}$$

com $\mathbf{W}$ e $\mathbf{V}$ como exposto acima.

A deviance do modelo utilizando a ligação *probit* será, considerando

$$\mu_i = \Phi(\eta_i),$$

tal que

$$\boldsymbol{\theta}_i = log\left(\frac{\mu_i}{1-\mu_i}\right) = log\left(\frac{\Phi(\eta_i)}{1-\Phi(\eta_i)}\right),$$

$$\hat{\boldsymbol{\theta}}_i = log\left(\frac{\hat{\mu}_i}{1-\hat{\mu}_i}\right) = log\left(\frac{\Phi(\hat{\eta}_i)}{1-\Phi(\hat{\eta}_i)}\right), \tilde{\boldsymbol{\theta}}_i = log\left(\frac{y_i/n}{1-y_i/n}\right)$$

$$b(\boldsymbol{\theta}_i) = -log(1-\mu_i),$$

$$b(\hat{\boldsymbol{\theta}}_i) = -log(1-\Phi(\hat{\eta}_i)), b(\tilde{\boldsymbol{\theta}}_i) = -log(1-y_i/n)$$

e

$$d_i = 2\sum_i^n \left\{ y_i(\tilde{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_i) + b(\hat{\boldsymbol{\theta}}_i) - b(\tilde{\boldsymbol{\theta}}_i) \right\}.$$

Portanto,

$$d_i = 2\sum_i^n \left\{ y_i log\left(\frac{y_i}{n\Phi(\hat{\eta}_i)}\right) + (m-y_i) log\left(\frac{1-y_i/n}{1-\Phi(\hat{\eta}_i)}\right) \right\}$$

para $0 < y_i < m$.

Se $y_i = 0, d_i = -2m\sum_i^n log(1-\Phi(\hat{\eta}_i))$ e se $y_i = m, d_i = -2m\sum_i^n log(\Phi(\hat{\eta}_i))$.

A ligação complemento log-log tem a seguinte forma

$$\eta_i = log(-log(1-\mu_i))$$

O mesmo ocorre para a ligação complemento log-log, a diferença na função escore e na matriz esperada de informação de Fisher será na matriz de pesos $\mathbf{W}$ e a matriz $\mathbf{V}$. Considere

$$e^{\eta_i} = -log(1-\mu_i)$$

e assim,

$$e^{-e^{\eta_i}} = 1-\mu_i.$$

Consequentemente,

$$\mu_i = 1 - e^{-e^{\eta_i}},$$

$$V_i = \mu_i(1-\mu_i) = (1-e^{-e^{\eta_i}})(1-1+e^{-e^{\eta_i}}) = e^{-e^{\eta_i}}(1-e^{-e^{\eta_i}}),$$

$$\frac{d\mu_i}{d\eta_i} = (e^{-e^{\eta_i}})(e^{\eta_i})$$

e

$$w_i = \frac{(e^{-e^{\eta_i}}e^{\eta_i})^2}{(1-e^{-e^{\eta_i}})(e^{-e^{\eta_i}})} = -\frac{(e^{\eta_i})^2 e^{-e^{\eta_i}}}{1-e^{-e^{\eta_i}}}.$$

A mesma definição para a função escore e matriz esperada de informação de Fisher se aplica a função de ligação complemento log-log, somente as matrizes $\mathbf{W}$ e $\mathbf{V}$ que serão definidas como mostrado acima.

Para a deviance com ligação complemento log-log, teremos

$$\mu_i = 1 - e^{-e^{\eta_i}},$$

$$\boldsymbol{\theta}_i = log\left(\frac{\mu_i}{1-\mu_i}\right), \tilde{\boldsymbol{\theta}}_i = log\left(\frac{y_i/n}{1-y_i/n}\right), \hat{\boldsymbol{\theta}}_i = log\left(\frac{1-e^{-e^{\hat{\eta}_i}}}{e^{-e^{\hat{\eta}_i}}}\right)$$

$$b(\boldsymbol{\theta}_i) = -log(1-\mu_i), b(\tilde{\boldsymbol{\theta}}_i) = -log(1-y_i/n), b(\hat{\boldsymbol{\theta}}_i) = -log(e^{-e^{\hat{\eta}_i}})$$

e

$$d_i = 2\sum_{i=1}^{n}\left\{y_i log\left(\frac{y_i}{n(1-e^{-e^{\hat{\eta}_i}})}\right) + (n-y_i)log\left(\frac{1-y_i/n}{e^{-e^{\hat{\eta}_i}}}\right)\right\}$$

para $0 < y_i < m$.

Se $y_i = 0$, $d_i = -2m\sum_{i=1}^{n} log(e^{-e^{\hat{\eta}_i}})$ e se $y_i = m$, $d_i = -2m\sum_{i=1}^{n} log(1-e^{-e^{\hat{\eta}_i}})$

4.5 INTERPRETAÇÃO PARA O MODELO BINOMIAL SEMIPARAMÉTRICO

Uma das vantagens do modelo semiparamétrico é que a interpretação para a razão de chances (OR) continuam as mesmas quando trabalhamos com a função de ligação *logit*.

O mesmo não ocorre para as funções de ligação *probit* e complemento log-log, onde mesmo no modelo paramétrico não é possível obter o valor da razão de chances.

Temos que

$$\hat{\boldsymbol{\mu}}(\mathbf{x},\mathbf{g}) = \frac{e^{\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{N}\hat{\mathbf{g}}}}{1+e^{\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{N}\hat{\mathbf{g}}}}, \hat{\boldsymbol{\mu}}(\mathbf{x},\mathbf{g}) = (1-e^{-e^{\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{N}\hat{\mathbf{g}}}}), \hat{\boldsymbol{\mu}}(\mathbf{x},\mathbf{g}) = \Phi(\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{N}\hat{\mathbf{g}})$$

será a probabilidade estimada de ocorrência referente as variáveis explicativas $\mathbf{x}$ e $\mathbf{g}$ para as funções de ligação *logit*, complemento log-log e *probit*, respectivamente.

Para a ligação *logit*, a chance de ocorrência estimada referente as variáveis explicativas $\mathbf{x}$ e $\mathbf{g}$ será definida por

$$\frac{\hat{\boldsymbol{\mu}}(\mathbf{x},\mathbf{g})}{1-\hat{\boldsymbol{\mu}}(\mathbf{x},\mathbf{g})} = e^{\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{N}\hat{\mathbf{g}}}.$$

E por fim, a razão de chances para uma unidade de mudança na k-ésima variável explicativa do modelo usando função de ligação *logit* será

$$OR = \frac{\frac{\hat{\boldsymbol{\mu}}(\mathbf{x}+\mathbf{1})}{1-\hat{\boldsymbol{\mu}}(\mathbf{x}+\mathbf{1})}}{\frac{\hat{\boldsymbol{\mu}}(\mathbf{x})}{1-\hat{\boldsymbol{\mu}}(\mathbf{x})}} = \frac{e^{\hat{\boldsymbol{\beta}}_1(x_{i1})+\ldots+\hat{\boldsymbol{\beta}}_k(x_{ik}+1)+\ldots+\hat{\boldsymbol{\beta}}_p x_{ip}+N\hat{g}}}{e^{\hat{\boldsymbol{\beta}}_1(x_{i1})+\ldots\hat{\boldsymbol{\beta}}_p x_{ip}+N\hat{g}}} = e^{\hat{\boldsymbol{\beta}}_k}$$

Quando temos uma variável explicativa dicotômica, $e^{\hat{\boldsymbol{\beta}}_k}$ é o estimador de razão de chances de ter o resultado para determinado grupo comparado com o outro grupo.

5 APLICAÇÃO

Nesta seção, iremos realizar simulações e uma aplicação a dados reais para o modelo binomial semiparamétrico. Utilizamos o algoritmo feito durante o trabalho para estimação dos parâmetros e da curva. Este algoritmo envolve splines cúbicos, matriz de informação esperada de Fisher, função escore, Newton-Raphson e outros métodos citados no decorrer deste trabalho. Tudo sendo realizado no software R Team (2017).

5.1 SIMULAÇÕES COM FUNÇÕES DE LIGAÇÃO LOGIT, COMPLEMENTO LOG-LOG E PROBIT

Considere os três modelos simulados

$$log\left(\frac{\mu}{1-\mu}\right) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Ng},$$

$$\eta = log(-log(1-\mu))$$

e

$$\Phi^{-1}(\mu) = \eta,$$

em que

-$\boldsymbol{\beta}_1 = -1$ e $\boldsymbol{\beta}_2 = 1$,

-t_1: uma sequência de 50 valores no intervalo de (0,4π),

-t: vetor de tamanho 100 onde os elementos são os valores do vetor t_1 repetidos uma vez, ou seja, uma junção de t_1 com t_1. Estes valores estão ordenados,

- $g(t) = sin(t)$,

-$n = 100, 200$ e 500,

-$x_1 \sim$ N(0;0,25) e $x_2 \sim$ N(0,1), tal que a matriz $\mathbf{X}$ tem dimensão 100x2,

-$\eta = \mathbf{X}\boldsymbol{\beta} + g(t)$,

-$\mu = \frac{e^{\eta}}{1+e^{\eta}}$,

-$m = 20$ (número de ensaios),

-$\mathbf{Y_1}$: número de sucessos em m ensaios, sendo que a probabilidade de sucesso em cada ensaio é igual a μ,

-$\mathbf{Y}^* = \frac{\mathbf{Y_1}}{m}$: proporção de sucessos em m ensaios,

-Intervalo para α: 1 a 5 e

- Em Eilers e Marx consideramos 14 subintervalos e grau do polinômio igual a 3.

Então, obtemos as estimativas que estão nas Tabela 3, 4 e 5, as curvas estimadas que estão nas Figuras 5, 6 e 7 , com o auxílio do método de Green e Silverman (1994), utilizando cada uma das três funções de ligação.

Tabela 3 – Média, Média dos Desvios-Padrão (SD) e Desvios-Padrão dos estimadores dos parâmetros (SDP) do modelo de regressão logística semiparamétrico (Green e Silverman) simulado com 100 replicações.

	Verdadeiro	n=100			n=200			n=500		
Parâmetro	Valor	Média	SD	SDP	Média	SD	SDP	Média	SD	SDP
$\boldsymbol{\beta}_1$	-1	-0,98	0,06	0,07	-0,99	0,05	0,05	-0,99	0,03	0,03
$\boldsymbol{\beta}_2$	1	0,94	0,12	0,13	1,00	0,07	0,08	1,00	0,05	0,04
α	-	4,99		10^{-5}	4,99		10^{-5}	4,99		10^{-5}

Tabela 4 – Média, Média dos Desvios-Padrão (SD) e Desvios-Padrão dos estimadores dos parâmetros (SDP) do modelo binomial semiparamétrico simulado com 100 replicações utilizando função de ligação complemento log-log(Green e Silverman).

	Verdadeiro	n=100			n=200			n=500		
Parâmetro	Valor	Média	SD	SDP	Média	SD	SDP	Média	SD	SDP
$\boldsymbol{\beta}_1$	-1	-0,98	0,05	0,05	-0,99	0,04	0,04	-0,99	0,02	0,02
$\boldsymbol{\beta}_2$	1	0,95	0,08	0,08	0,99	0,06	0,06	0,99	0,03	0,03
α	-	4,78		0,54	4,82		0,47	3,99		0,93

Tabela 5 – Média, Média dos Desvios-Padrão (SD) e Desvios-Padrão dos estimadores dos parâmetros (SDP) do modelo binomial semiparamétrico simulado com 100 replicações utilizando função de ligação probit(Green e Silverman).

	True	n=100			n=200			n=500		
Parameter	Value	Média	SD	SDP	Média	SD	SDP	Média	SD	SDP
$\boldsymbol{\beta}_1$	-1	-0,98	0,04	0,05	-0,99	0,03	0,03	-0,99	0,02	0,02
$\boldsymbol{\beta}_2$	1	0,99	0,08	0,07	0,99	0,05	0,06	0,99	0,03	0,03
α	-	4,99		10^{-5}	4,99		10^{-5}	4,99		10^{-5}

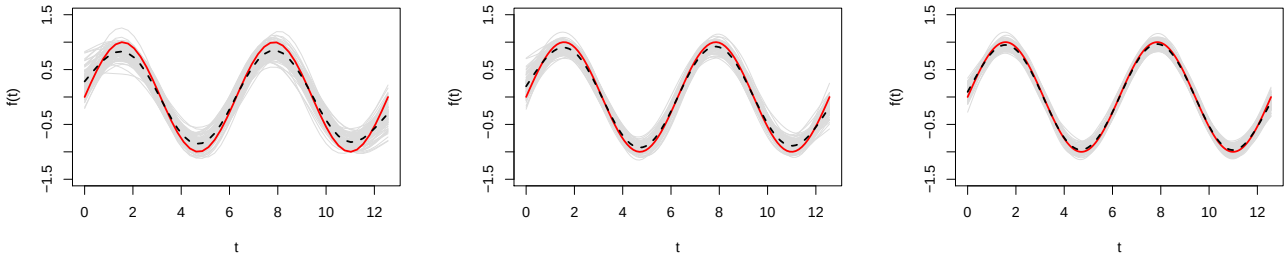


Figura 5 – Gráficos para a componente não-paramétrica do modelo binomial, utilizando função de ligação logit(Green e Silverman) com 100 replicações. Curvas estimadas (linhas cinza), curva verdadeira (linha vermelha) e média das curvas estimadas (linha preta tracejada): a) n=100 b) n=200 e c) n=500.

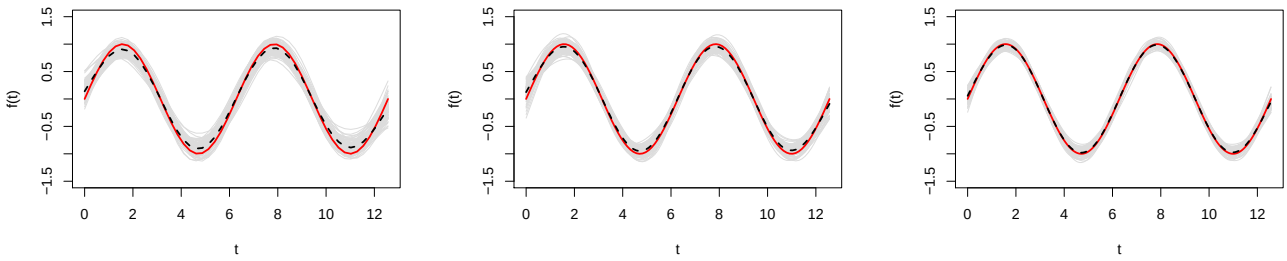


Figura 6 – Gráficos para a componente não-paramétrica do modelo binomial, utilizando função de ligação complemento log-log(Green e Silverman) com 100 replicações. Curvas estimadas (linhas cinza), curva verdadeira (linha vermelha) e média das curvas estimadas (linha preta tracejada): a) n=100 b) n=200 e c) n=500.

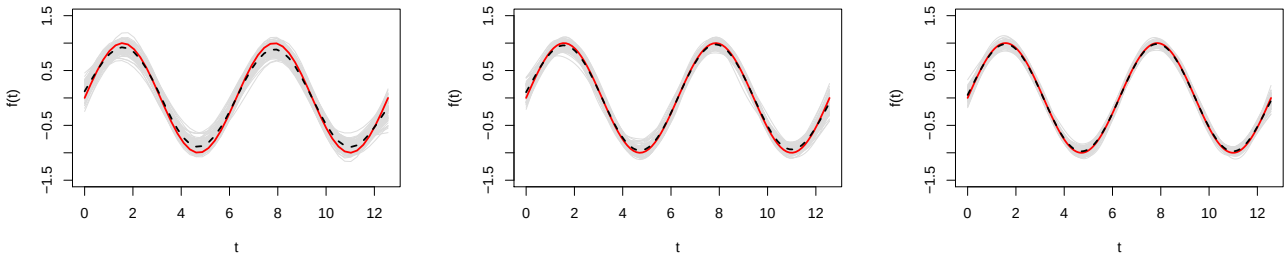


Figura 7 – Gráficos para a componente não-paramétrica do modelo binomial, utilizando função de ligação probit(Green e Silverman) com 100 replicações. Curvas estimadas (linhas cinza), curva verdadeira (linha vermelha) e média das curvas estimadas (linha preta tracejada): a) n=100 b) n=200 e c) n=500.

Além disso, obtemos as estimativas que estão nas Tabelas 6, 7 e 8, as curvas estimadas que estão nas Figuras 8, 9 e 10 pelo método de Eilers e Marx utilizando cada uma das três funções de ligação.

Tabela 6 – Média, Média dos Desvios-Padrão (SD) e Desvios-Padrão dos estimadores dos parâmetros (SDP) do modelo de regressão logística semiparamétrico(Eilers e Marx) simulado com 100 replicações.

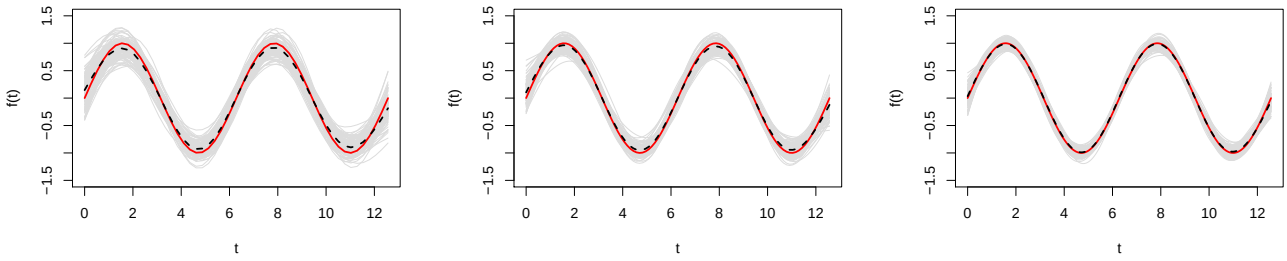
	Verdadeiro	n=100			n=200			n=500		
Parâmetro	Valor	Média	SD	SDP	Média	SD	SDP	Média	SD	SDP
$\boldsymbol{\beta}_1$	-1	-0,99	0,06	0,06	-0,99	0,04	0,04	-0,99	0,03	0,03
$\boldsymbol{\beta}_2$	1	0,99	0,11	0,11	0,99	0,08	0,07	0,99	0,05	0,05
α	-	3,25		0,89	3,31		0,63	4,87		0,56

Tabela 7 – Média, Média dos Desvios-Padrão (SD) e Desvios-Padrão dos estimadores dos parâmetros (SDP) do modelo binomial semiparamétrico com função de ligação complemento log-log(Eilers e Marx) simulado com 100 replicações.

	Verdadeiro	n=100			n=200			n=500		
Parâmetro	Valor	Média	SD	SDP	Média	SD	SDP	Média	SD	SDP
$\boldsymbol{\beta}_1$	-1	-0,98	0,06	0,05	-0,99	0,04	0,04	-0,99	0,02	0,02
$\boldsymbol{\beta}_2$	1	0,96	0,08	0,08	1,00	0,06	0,06	0,99	0,03	0,03
α	-	4,28		0,94	4,92		0,32	4,63		0,64

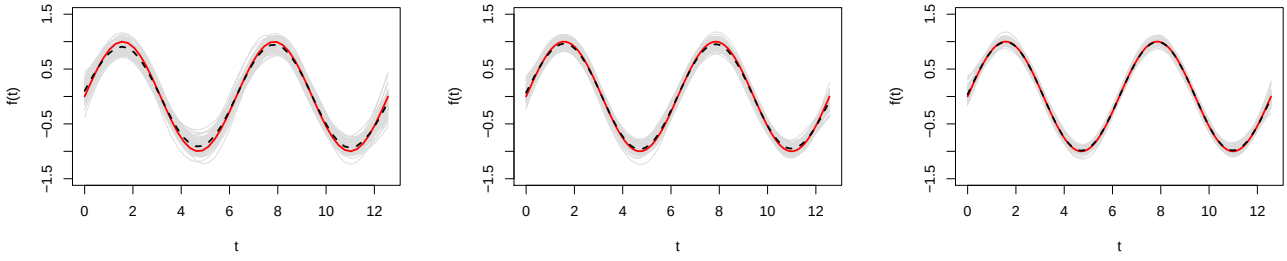
Tabela 8 – Média, Média dos Desvios-Padrão (SD) e Desvios-Padrão dos estimadores dos parâmetros (SDP) do modelo binomial semiparamétrico simulado usando função de ligação probit(Eilers e Marx) com 100 replicações.

	Verdadeiro	n=100			n=200			n=500		
Parâmetro	Valor	Média	SD	SDP	Média	SD	SDP	Média	SD	SDP
$\boldsymbol{\beta}_1$	-1	-0,99	0,05	0,05	-0,99	0,03	0,03	-0,99	0,02	0,02
$\boldsymbol{\beta}_2$	1	1,02	0,07	0,07	0,99	0,05	0,05	0,99	0,03	0,03
α	-	3,93		0,87	4,46		0,82	4,61		0,69



(a) **(b)** **(c)**

Figura 8 – Gráficos para a componente não-paramétrica do modelo binomial, utilizando função de ligação logit(Eilers e Marx) com 100 replicações. Curvas estimadas (linhas cinza), curva verdadeira (linha vermelha) e média das curvas estimadas (linha preta tracejada): a) n=100 b) n=200 e c) n=500.



(a) **(b)** **(c)**

Figura 9 – Gráficos para a componente não-paramétrica do modelo binomial, utilizando função de ligação complemento log-log(Eilers e Marx) com 100 replicações. Curvas estimadas (linhas cinza), curva verdadeira (linha vermelha) e média das curvas estimadas (linha preta tracejada): a) n=100 b) n=200 e c) n=500.

Quando trabalhamos com a variável resposta sendo o número de sucessos em m ensaios e este m sendo maior que um, conseguimos retirar do modelo os seus resíduos. A Figura 11 evidencia os resíduos não-paramétricos do modelo binomial semiparamétrico com função de ligação logit, utilizando o método de Green e Silverman. Os dados simulados do gráfico foram os mesmos trabalhados nas simulações (usando n=100). Podemos observar que a curva estimada acompanha os resíduos.

O intuito principal das simulações mostradas nesta seção era o de comprovar a validade dos cálculos feitos para a matriz esperada de informação de Fisher, função escore e validação cruzada, propostos para auxiliar na estimação dos parâmetros do modelo binomial semiparamétrico, considerando cada uma das funções de ligação e os métodos para se tratar a curva suave. Note que, obtemos estimativas precisas ao que foi simulado e ajustes

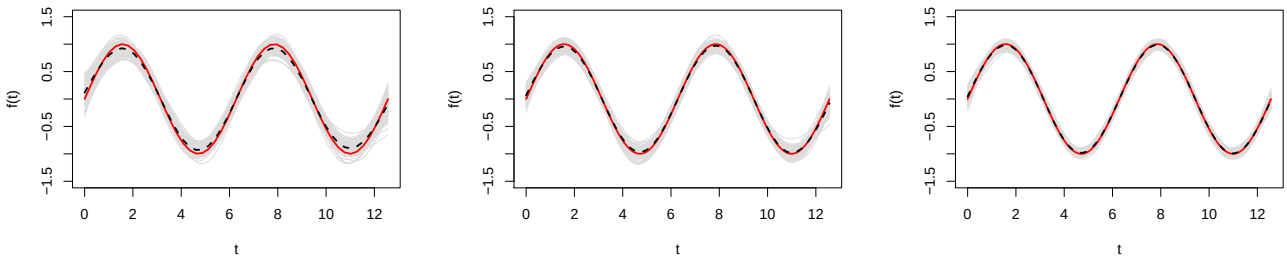


Figura 10 – Gráficos para a componente não-paramétrica do modelo binomial, utilizando função de ligação probit(Eilers e Marx) com 100 replicações. Curvas estimadas (linhas cinza), curva verdadeira (linha vermelha) e média das curvas estimadas (linha preta tracejada): a) n=100 b) n=200 e c) n=500.

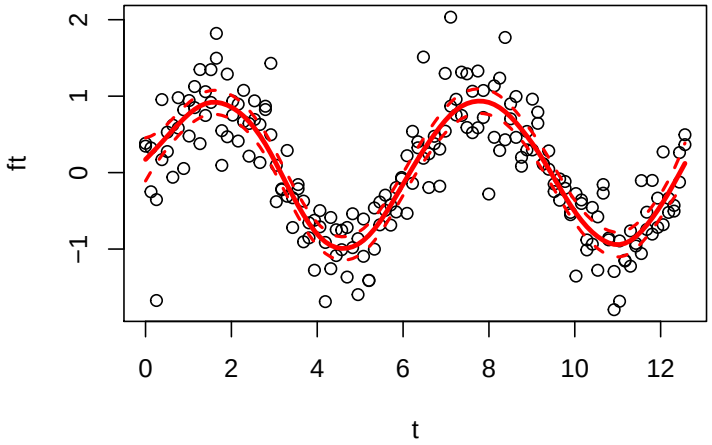


Figura 11 – Gráfico da componente não-paramétrica do modelo binomial semiparamétrico (função de ligação logit). Em vermelho curva estimada, os pontos são referentes aos resíduos não-paramétricos e as bandas de confiança, para a curva estimada, em vermelho tracejado.

de curvas muito satisfatórios, assemelhando-se a curva verdadeira. Então, podemos afirmar que nosso algoritmo está retornando valores iguais a nossos valores de entrada(simulados).

5.2 APLICAÇÃO A DADOS REAIS

Dinse e Lagakos[3] (1984) relatam uma análise de modelo binomial de alguns dados de um estudo do Programa Nacional de Toxicologia dos EUA sobre retardadores de chama. Os dados sobre ratos machos e fêmeas expostos a várias doses de uma mistura de bifenatos polibromados, conhecidos como Firemaster FF-1. A variável resposta consiste em uma variável binária, $\mathbf{Y}$, indicando presença ou ausência de uma lesão, hiperplasia

do ducto biliar, na morte de cada animal, causada pela exposição a essa substância presente nos retardadores de chama. Foram analisados 319 ratos. Consideramos quatro variáveis explicativas: dose, x_1, peso inicial, x_2, posição da gaiola (altura acima do chão), $\mathbf{g}$, sexo, x_3 (0:fêmea e 1:macho) e t é o tempo de sobrevivência em semanas do animal. Trabalhamos com as variáveis x_1, x_2 e x_3 de forma paramétrica e a variável peso, $\mathbf{g}$, de forma não-paramétrica.

Utilizando o mesmo algoritmo trabalhado nas simulações, chegamos às estimativas para o vetor de parâmetros $\boldsymbol{\beta}$ e as curvas estimadas utilizando os três tipos de funções de ligação, usando Green e Silverman[7] (1994) e Eilers e Marx[4] (1996). Tabelas 9, 10 e 11 e Figura 11, são referentes a Green e Silverman[7] (1994). As Tabelas 12, 13 e 14, e Figura 14 referentes a Eilers e Marx[4] (1996).

Tabela 9 – Estimativas para os dados reais com regressão logística semiparamétrica utilizando Green e Silverman

	Estimativas	Erro Padrão
$\boldsymbol{\beta}_1$ (posição)	0,13	0,10
$\boldsymbol{\beta}_2$ (dose)	0,11	0,09
$\boldsymbol{\beta}_3$ (sexo)	0,82	0,35
α	799,99	-

Tabela 10 – Estimativas para os dados reais com modelo binomial semiparamétrico com função de ligação complemento log-log, utilizando Green e Silverman.

	Estimativas	Erro Padrão
$\boldsymbol{\beta}_1$ (posição)	0,09	0,08
$\boldsymbol{\beta}_2$ (dose)	0,07	0,08
$\boldsymbol{\beta}_3$ (sexo)	0,69	0,31
α	799.999	-

Tabela 11 – Estimativas para os dados reais para modelo binomial semiparamétrico com função de ligação *probit*, utilizando Green e Silverman.

	Estimativas	Erro Padrão
$\boldsymbol{\beta}_1$ (posição)	0,08	0,06
$\boldsymbol{\beta}_2$ (dose)	0,07	0,05
$\boldsymbol{\beta}_3$ (sexo)	0,51	0,19
α	750,00	-

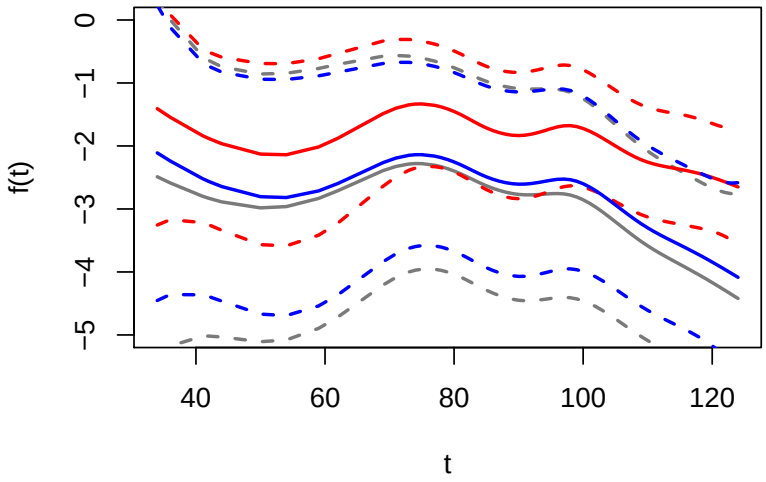


Figura 12 – Curvas estimadas para os dados reais no modelo binomial semiparamétrico. Curva estimada em cinza pela função de ligação *logit*, em vermelho função de ligação *probit* e em azul função de ligação complemento log-log e suas respectivas bandas de confiança em linhas tracejadas. Utilizando Green e Silverman

Tabela 12 – Estimativas para os dados reais para modelo binomial semiparamétrico com função de ligação *probit*, utilizando Eilers e Marx.

	Estimativas	Erro Padrão
$\boldsymbol{\beta}_1$ (posição)	0,08	0,06
$\boldsymbol{\beta}_2$ (dose)	0,07	0,05
$\boldsymbol{\beta}_3$ (sexo)	0,51	0,198
α	1,27	-

Tabela 13 – Estimativas para dados reais para modelo binomial semiparamétrico com função de ligação *logit*, utilizando Eilers e Marx.

	Estimativas	Erro Padrão
$\boldsymbol{\beta}_1$ (posição)	0,13	0,10
$\boldsymbol{\beta}_2$ (dose)	0,11	0,09
$\boldsymbol{\beta}_3$ (sexo)	0,83	0,35
α	1	-

Tabela 14 – Estimativas para dados reais com modelo binomial semiparamétrico com função de ligação complemento log-log, utilizando Eilers e Marx.

	Estimativas	Erro Padrão
$\boldsymbol{\beta}_1$ (posição)	0,09	0,08
$\boldsymbol{\beta}_2$ (dose)	0,07	0,07
$\boldsymbol{\beta}_3$ (sexo)	0,70	0,31
α	1	-

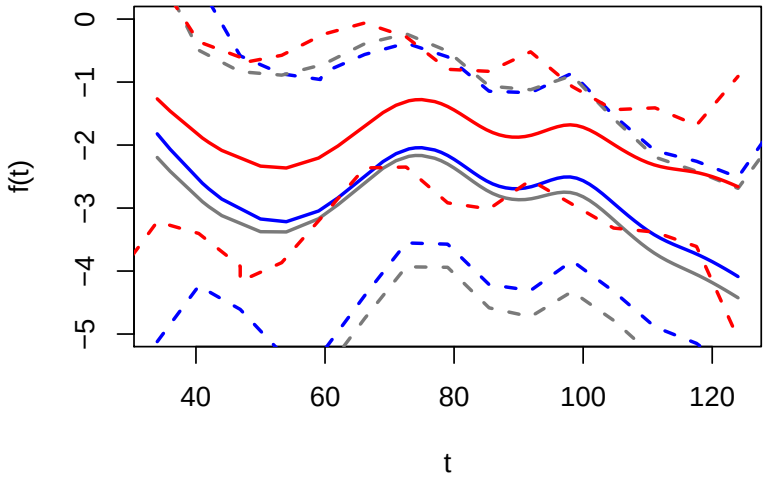


Figura 13 – Curvas estimadas para aplicação modelo binomial semiparamétrico. Curva estimada em cinza pela função de ligação *logit*, em vermelho função de ligação probit e em azul função de ligação complemento log-log e suas respectivas bandas de confiança em linhas tracejadas, utilizando Eilers e Marx.

Tabela 15 – AIC e BIC para cada método utilizado na aplicação aos dados reais

	Green		Eilers	
Função de ligação	AIC	BIC	AIC	BIC
logit	335,04	376,46	333,51	374,93
probit	331,52	372,94	330,53	371,95
clog-log	335,08	376,49	333,32	374,74

O objetivo principal desta aplicação a dados reais não é a interpretação dos parâmetros e consequentemente a conclusão da relação entre as variáveis explicativas paramétricas e resposta, por isso não faremos as interpretações dos mesmos, mas essa interpretação é obtido seguindo o que foi exposto na seção 4.5.

Como dito anteriormente, testamos três critérios para trabalhar com a validação cruzada para o modelo binomial semiparamétrico (Seção 4.3), mas nos três critérios obtemos resultados parecidos. A estimativa para o α tendia sempre ao limite superior do intervalo de entrada, então optamos por fazer as simulações e aplicação a dados reais com a validação cruzada que utiliza a matriz $\mathbf{A}$, devido a maior simplicidade e rapidez do algoritmo. Este comportamento de α muda quando trabalhamos com o método de Eilers e Marx[4] (1996), neste método o α em média não atinge o limite superior do intervalo de entrada. Na aplicação aos dados reais, consideramos o intervalo para α de 750 a 800 em Green e Silverman[7] (1994) e de 1 a 5 em Eilers e Marx[4] (1996). A explicação para esta diferença tão grande entre os valores de α para os dois métodos é a mesma utilizada nas simulações. Eilers e Marx[4] (1996) absorvem mais informações sobre a curva devido a maneira de escolha de seus nós, assim isto impacta no valor de α. Como já possuem muitas características da curva não necessitam de um valor de α alto para suavizá-la.

Na aplicação aos dados reais observamos que utilizando as funções de ligação *probit* e complemento log-log, obtemos estimativas próximas, mas quando analisamos o ajuste da curva as funções de ligação *logit* e complemento log-log que se assemelham . Isto ficou evidente tanto trabalhando com Green e Silverman[7] (1994) como Eilers e Marx[7] (1996). Destacando que, mesmo as estimativas usando as funções de ligação *probit* e complemento log-log se assemelharem mais, não significa que elas se distanciam das estimativas usando função de ligação *logit*.

Analisando os valores dos AIC e BIC na Tabela 15, vemos que de acordo com estes critérios o melhor modelo seria o modelo binomial semiparamétrico utilizando a função de ligação *probit*, pois foi a que obteve os menores valores de ambos, ou seja, estaríamos perdendo menos informação com esta função de ligação. *Logit* e complemento log-log possuem valores muito próximos para estes critérios.

No cálculo dos graus de liberdade para a parte não-paramétrica do modelo usamos o traço da matriz $\mathbf{N}(\mathbf{N}^T\mathbf{N} + \alpha\mathbf{K})^{-1}\mathbf{N}^T$ e para a parte paramétrica do modelo usamos o tamanho do vetor de parâmetros $\boldsymbol{\beta}$. Assim, o número de graus de liberdade total será a soma dos graus de liberdade da parte paramétrica e não-paramétrica do modelo.

Trabalhando com dados reais, onde o número de tentativas é igual a 1 (m=1) não é possível obter do modelo os seus respectivos resíduos.

6 CONCLUSÕES

Um dos objetivos que este trabalho se propôs a cumprir, ao se deparar com a dificuldade ou até mesmo ausência de referências sobre o assunto na literatura, foi dar toda base possível para o entendimento do modelo binomial semiparamétrico. Por isso, com este trabalho, contribuímos de forma detalhada com todas as etapas de construção deste modelo, desde o cálculo da matriz esperada de informação de Fisher, função escore, validação cruzada, deviance, entre outras, até a estimação dos parâmetros e curva suave, isto para cada uma das três funções de ligação e os dois métodos para estimação da curva suave. Também damos detalhes para implementação de códigos que podem ser reproduzidos.

Um dos trabalhos futuros é continuar trabalhando com Modelos Lineares Generalizados de forma semiparamétrica, mas com a variável resposta seguindo outras distribuições, como Poisson, Normal, Normal Inversa e Gama. E tendo a possibilidade de acrescentar mais uma variável não-paramétrica, pois em nosso trabalho utilizamos somente uma variável não-paramétrica e a proposta seria trabalhar com mais de uma.

Uma ideia que surgiu durante o trabalho foi produzir um algoritmo onde o usuário pudesse escolher a localização dos nós. Isto ajudaria em determinadas curvas que em algumas regiões existem a necessidade de se ter uma maior quantidade de nós e em outras uma menor. Este algoritmo foi feito, mas vendo a necessidade de mais estudos e certificados de sua comprovação, não o colocamos neste trabalho.

REFERÊNCIAS

[1] CHARNET, Reinaldo et al. *Análise de modelos de regressão linear com aplicações*. Campinas, São Paulo, Unicamp, 356p, 1999.

[2] DE BOOR, Carl et al. A practical guide to splines. *New York: Springer-Verlag*, 1978.

[3] DINSE, Gregg E.; LAGAKOS, S. W. Regression analysis of tumour prevalence data. *Wiley for the Royal Statistical Society*, p. 79-80, 1984.

[4] EILERS, Paul HC; MARX, Brian D. Flexible smoothing with B-splines and penalties. *Statistical science*, p. 89-102, 1996.

[5] EUBANK, Randall L. Nonparametric regression and spline smoothing. *CRC press*, 1999.

[6] GOLUB, Gene H.; HEATH, Michael; WAHBA, Grace. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, v. 21, n. 2, p. 215-223, 1979.

[7] GREEN, Peter J.; SILVERMAN, Bernard W. *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall, 1994.

[8] HASTIE, Trevor; TIBSHIRANI, Robert. Generalized additive models: some applications. *Journal of the American Statistical Association*, v. 82, n. 398, p. 371-386, 1987.

[9] O'SULLIVAN, Finbarr; YANDELL, Brian S.; RAYNOR JR, William J. Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association*, v. 81, p. 96-103, 1986.

[10] PAULA, Gilberto Alvarenga. *Modelos de regressão: com apoio computacional*. São Paulo: IME-USP, 2004.

[11] R Development Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing,Vienna, Austria. ISBN 3-900051-07-0,URL http://www.R-project.org/.