
A distribuição normal-valor extremo generalizado
para a modelagem de dados limitados no
intervalo unitário $(0,1)$

Yury Rojas Benites

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

Yury Rojas Benites

A distribuição normal-valor extremo generalizado para a modelagem de dados limitados no intervalo unitário $(0,1)$

Dissertação apresentada ao Departamento de Estatística – DEs-UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestra em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Dr. Vicente Garibay Cancho

São Carlos
Julho de 2019

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

Yury Rojas Benites

The normal-generalized extreme value distribution for the modeling of data restricted in the unit interval $(0,1)$

Master dissertation submitted to the Department of Statistics - DEs-UFSCar and to the Institute of Mathematics and Computer Sciences - ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics.

Advisor: Prof. Dr. Vicente Garibay Cancho

São Carlos
July 2019

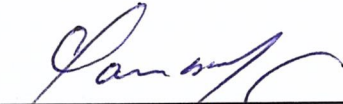


UNIVERSIDADE FEDERAL DE SÃO CARLOS

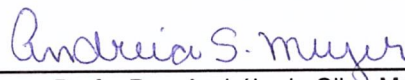
Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

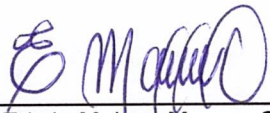
Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Yury Rojas Benites, realizada em 28/06/2019:



Prof. Dr. Vicente Garibay Cancho
USP



Profa. Dra. Andréia da Silva Meyer
UNESP



Prof. Dr. Edwin Moisés Marcos Ortega
ESALQ/USP

*Dedico este trabalho primeiramente a Deus, meu guia, meu socorro, meu tudo.
A minha mãe e aos meus irmãos, força e alegria do meu coração.
E a aquelas pessoas que acreditaram em mim neste caminho, reconforto da minha vida.*

AGRADECIMENTOS

Primeiramente agradeço a Deus, por nunca me deixar sozinha neste caminho, por todas as coisas que tive que passar para ser melhor, e por todas as bênçãos que ele me deu.

Aos meus pais, Isabel e Máximo, por me darem a vida, os estudos e sobre tudo por me darem amor. Agradeço de modo especial a minha mãe por ser meu exemplo de vida, por todos seus esforços, e por me ensinar a acreditar que tudo é possível, obrigada mãe!. Aos meus irmãos por seu carinho e apoio, de modo especial a meu irmão Arturo, porque houve um momento que eu quis desistir de tudo, então foi ele que me deu palavras de ânimo para continuar neste caminho e fazer valer a pena todo sacrifício feito.

Ao professor Vicente Garibay Cancho, por sua orientação, sua paciência, por seus conselhos, sua confiança, e sobre tudo por ser como um pai para mim, algo que sempre estarei muito agradecida.

A minha amiga querida Elizbeth, por acreditar em mim e ser o anjo que me acompanhou neste caminho.

A meu colega Caio, por sua colaboração com o texto do trabalho.

A todos meus amigos e amigas, por seus conselhos, amizade e carinho.

Aos professores, funcionários do PIPGES USP-UFSCar, assim como ao pessoal da biblioteca Prof. Achille Bassi do ICMC, pelos conhecimentos partilhados, o excelente serviço e a ajuda brindada.

Aos membros da banca, da qualificação e da defesa, por todas suas sugestões e conselhos que são de enriquecimento do trabalho.

Ao Brasil, pela oportunidade.

E finalmente agradeço a CAPES, porque o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

*“O que é impossível para os homens,
é possível para Deus.”*

(Jesus)

Lc 18:27

RESUMO

BENITES, Y. R. **A distribuição normal-valor extremo generalizado para a modelagem de dados limitados no intervalo unitário** $(0, 1)$. 2019. 97 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Neste trabalho é introduzido um novo modelo estatístico para modelar dados limitados no intervalo contínuo $(0, 1)$. O modelo proposto é construído sob uma transformação de variáveis, onde a variável transformada é resultado da combinação de uma variável com distribuição normal padrão e a função de distribuição acumulada da distribuição valor extremo generalizado. Para o novo modelo são estudadas suas propriedades estruturais. A nova família é estendida para modelos de regressão, onde o modelo é reparametrizado na mediana da variável resposta e este conjuntamente com o parâmetro de dispersão são relacionados com covariáveis através de uma função de ligação. Procedimentos inferenciais são desenvolvidos desde uma perspectiva clássica e bayesiana. A inferência clássica baseia-se na teoria de máxima verossimilhança e a inferência bayesiana no método de Monte Carlo via cadeias de Markov. Além disso estudos de simulação são realizados para avaliar o desempenho das estimativas clássicas e bayesianas dos parâmetros do modelo, em que, os estimadores de máxima verossimilhança dos parâmetros do modelo satisfazem as propriedades assintóticas. Finalmente um conjunto de dados de câncer colorretal é considerado para mostrar a aplicabilidade do modelo, em que, em comparação com outros modelos, o modelo proposto é o que melhor ajusta este conjunto de dados.

Palavras-chave: Distribuição valor extremo generalizado, Estimador de máxima verossimilhança, Inferência bayesiana, Método MCMC.

ABSTRACT

BENITES, Y. R. **The normal-generalized extreme value distribution for the modeling of data restricted in the unit interval $(0, 1)$** . 2019. 97 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

In this research a new statistical model is introduced to model data restricted in the continuous interval $(0, 1)$. The proposed model is constructed under a transformation of variables, in which the transformed variable is the result of the combination of a variable with standard normal distribution and the cumulative distribution function of the generalized extreme value distribution. For the new model its structural properties are studied. The new family is extended to regression models, in which the model is reparametrized in the median of the response variable and together with the dispersion parameter are related to covariables through a link function. Inferential procedures are developed from a classical and Bayesian perspective. The classical inference is based on the theory of maximum likelihood, and the Bayesian inference is based on the Markov chain Monte Carlo method. In addition, simulation studies were performed to evaluate the performance of the classical and Bayesian estimates of the model parameters, in which the maximum likelihood estimators of the model parameters satisfy the asymptotic properties. Finally a set of colorectal cancer data is considered to show the applicability of the model, in which, compared to other models, the proposed model best fits this dataset.

Keywords: Generalized extreme value distribution, Maximum likelihood estimator, Bayesian inference, MCMC Method.

LISTA DE ILUSTRAÇÕES

Figura 1 – Função de densidade de probabilidade da distribuição VEG	26
Figura 2 – Função de distribuição acumulada da distribuição VEG	26
Figura 3 – Função de distribuição acumulada da distribuição VEG e da distribuição logística padrão	26
Figura 4 – Gráfico da função densidade de probabilidade da distribuição Normal-VEG com λ variando.	30
Figura 5 – Gráfico da função densidade de probabilidade da distribuição Normal-VEG com δ variando.	30
Figura 6 – Gráfico da função densidade de probabilidade da distribuição Normal-VEG, com γ variando.	31
Figura 7 – Coeficiente de assimetria da distribuição Normal-VEG com parâmetros: (a) $\delta = 2, \gamma = 0,5$, (b) $\delta = 2, \gamma = 0,2$, (c) $\delta = 4, \gamma = 0$, (d) $\delta = 2, \gamma = -0,2$	33
Figura 8 – Ajuste do Modelo Normal-VEG, do Modelo S_B Johnson, do Modelo GJS-t-Student, e do Modelo Beta aos dados	42
Figura 9 – Boxplot da taxa de sobrevivência de pacientes com câncer colorretal por sexo (lado esquerdo) e por raça (lado direito)	64
Figura 10 – QQ-plot de resíduos quantílicos aleatorizados para os modelos Normal-VEG e S_B Johnson	66
Figura 11 – Gráfico de traço para os parâmetros do modelo Normal-VEG-LL para os dados de câncer colorretal	79
Figura 12 – Densidade marginal a posteriori do parâmetro λ do modelo de regressão Normal-VEG-LL (lado esquerdo), o QQ-plot dos resíduos quantílicos aleatorizados normalizados a posteriori do modelo Normal-VEG-LL (lado direito) e do modelo de regressão S_B de Johnson com ligação log-log (lado central)	81
Figura 13 – Gráficos de índices das medidas ψ -divergência para os dados de câncer colorretal	82
Figura 14 – Gráfico das densidades a posteriori da mediana da taxa de sobrevivência para seis pacientes hipotéticos	83

LISTA DE TABELAS

Tabela 1 – Média e variância da Distribuição Normal-VEG obtidas numericamente para diferentes valores dos parâmetros.	32
Tabela 2 – Resultados da simulação do modelo Normal-VEG com parâmetros $\lambda = -0,2; 0; 0,2$, com $\gamma = 2$ e $\delta = 10$	37
Tabela 3 – Resultados da simulação do modelo Normal-VEG com $\lambda = -0,1; 0; 0,1$, com $\gamma = 0,1$ e $\delta = 2$	38
Tabela 4 – Resultados da simulação do modelo Normal-VEG com $\lambda = -0,5; 0; 0,5$, com $\gamma = -0,2$ e $\delta = 6$	39
Tabela 5 – Resultados da simulação do modelo Normal-VEG com $\lambda = -0,2; 0; 0,2$, com $\gamma = -0,1$ e $\delta = 1$	40
Tabela 6 – Medidas descritivas da taxa de incidência de Câncer Colorretal nos EUA . . .	41
Tabela 7 – Estimativas de máxima verossimilhança e erro padrão dos parâmetros dos modelos Normal-VEG, S_B de Johnson, GJS-t-student(4) e Beta para a taxa de incidência de câncer colorretal.	41
Tabela 8 – Critérios de comparação de modelos AIC e BIC para os modelos Normal-VEG, S_B de Johnson, GJS-t-Student(4) e Beta para a taxa de incidência de câncer colorretal.	42
Tabela 9 – Funções de ligação para a mediana e para o parâmetro de dispersão	46
Tabela 10 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG homocedástico com parâmetros $\beta_0 = 1, \beta_1 = -2, \delta = 2$ e $\lambda = 0,2$. . .	50
Tabela 11 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG homocedástico com parâmetros $\beta_0 = -1, \beta_1 = 6, \delta = 2$ e $\lambda = -0,2$. . .	51
Tabela 12 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG homocedástico com parâmetros $\beta_0 = -1, \beta_1 = 0,2, \delta = 10$ e $\lambda = -0,4$. . .	52
Tabela 13 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG homocedástico com parâmetros $\beta_0 = 1, \beta_1 = -0,2, \delta = 4$ e $\lambda = -0,6$. . .	53
Tabela 14 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = 1, \beta_1 = -2, \tau_0 = 2, \tau_1 = 1$ e $\lambda = -0,2$	56
Tabela 15 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = 1, \beta_1 = -2, \tau_0 = 2, \tau_1 = 1$ e $\lambda = -0,7$	57

Tabela 16 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = -4$, $\beta_1 = -2$, $\tau_0 = 2$, $\tau_1 = 1$ e $\lambda = 0,7$	58
Tabela 17 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$ e $\lambda = 0,4$	59
Tabela 18 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = -6$, $\beta_1 = 8$, $\tau_0 = 0,2$, $\tau_1 = 1$ e $\lambda = -0,2$	60
Tabela 19 – Resultados baseados em 1000 simulações do modelo de regressão heterocedástico com parâmetros $\beta_0 = -1$, $\beta_1 = 2$, $\tau_0 = 1$, $\tau_1 = 0,4$ e $\lambda = 0,5$	61
Tabela 20 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = -5$, $\beta_1 = -1$, $\tau_0 = 1$, $\tau_1 = 2$ e $\lambda = 0,1$	62
Tabela 21 – Medidas descritivas da taxa de sobrevivência das pessoas com câncer colorretal nos EUA	63
Tabela 22 – Critérios de comparação de modelos AIC e BIC do modelo Normal-VEG com diferentes funções de ligações para os dados de câncer colorretal	65
Tabela 23 – Estimativas de Máxima Verossimilhança dos parâmetros dos modelos de regressão Normal-VEG e S_B de Johnson.	65
Tabela 24 – Critério AIC e BIC para os modelos ajustados.	66
Tabela 25 – Estimativa da mediana da taxa de sobrevivência e o intervalo de 95% de confiança para os seis pacientes hipotéticos com câncer colorretal.	67
Tabela 26 – Estimativas bayesianas sob perda quadrática e perda absoluta dos parâmetros do modelo de regressão Normal-VEG, desvio padrão (DP), viés, a raiz do erro quadrático médio (REQM) e a probabilidade de cobertura (PC), baseados em 1000 simulações do modelo com parâmetros $\lambda = -0,4$, $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$	75
Tabela 27 – Estimativas bayesianas sob perda quadrática e perda absoluta dos parâmetros do modelo de regressão Normal-VEG, DP, viés, REQM e PC, baseados em 1000 simulações do modelo com parâmetros $\lambda = 0,4$, $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$	76
Tabela 28 – Estimativas bayesianas sob perda quadrática e perda absoluta dos parâmetros do modelo de regressão Normal-VEG, DP, viés, REQM e PC, baseados em 1000 simulações do modelo com parâmetros $\lambda = -0,1$, $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$	76

Tabela 29 – Estimativas bayesianas sob perda quadrática e perda absoluta dos parâmetros do modelo de regressão Normal-VEG, DP, viés, REQM e PC, baseados em 1000 simulações do modelo com parâmetros $\lambda = 0, 1$, $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$	77
Tabela 30 – Critérios de comparação de modelos DIC, EAIC, EBIC, e LPML do modelo de regressão Normal-VEG com diferentes funções de ligações para os dados de câncer colorretal	78
Tabela 31 – Estimativas e intervalo HPD (95%) dos parâmetros do modelo de regressão Normal-VEG e S_B de Johnson com função de ligação log-log para os dados de câncer colorretal	80
Tabela 32 – Critério DIC, EAIC, EBIC e LPML para os modelos de regressão Normal-VEG, S_B de Johnson, e GJS-t-Student(4) com função de ligação log-log para os dados de câncer colorretal	80
Tabela 33 – Estimativas da taxa de sobrevivência mediana e o intervalo HPD de 95% de credibilidade para os seis pacientes hipotéticos com câncer colorretal.	82
Tabela 34 – Medidas de ψ -divergência do ajuste do modelo de regressão Normal-VEG-LL para os dados de câncer colorretal	82
Tabela 35 – $MRS(\%)$ e seu correspondente intervalo HPD de 95% de credibilidade (LI:limite inferior, LS:limite superior) ajustando o modelo de regressão Normal-VEG aos dados de câncer colorretal sem as observações influentes .	83

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Distribuição Valor Extremo Generalizado (VEG)	25
1.2	Objetivos	27
1.3	Organização do trabalho	27
2	MODELO NORMAL-VEG	29
2.1	Modelo Normal-VEG	29
2.2	Inferência	32
2.3	Estudo de simulação	35
2.4	Aplicação	41
2.5	Conclusão	42
3	MODELO DE REGRESSÃO NORMAL-VEG	43
3.1	O modelo Normal-VEG parametrizado na mediana	43
3.1.1	<i>Parâmetro de dispersão</i>	44
3.2	Modelo de Regressão Normal-VEG	45
3.3	Inferência	46
3.3.1	<i>Função Escore</i>	46
3.3.2	<i>Teste da razão de verossimilhança</i>	48
3.4	Estudo de simulação para o modelo de regressão Normal-VEG homocedástico	48
3.5	Estudo de simulação para o modelo de regressão Normal-VEG heterocedásticos	54
3.6	Aplicação	63
3.7	Conclusão	67
4	MODELO DE REGRESSÃO NORMAL-VEG: ENFOQUE BAYESIANO	69
4.1	Modelo de regressão Normal-VEG	69
4.2	Inferência Bayesiana	70
4.3	Critério de comparação de modelos	71
4.4	Análise de diagnóstico bayesiano	72
4.5	Estudo de simulação	74
4.6	Aplicação	77

4.7	Conclusão	84
5	CONSIDERAÇÕES FINAIS E PROPOSTAS FUTURAS	85
	REFERÊNCIAS	87
ANEXO A	FUNÇÃO ESCORE NORMAL-VEG	91
ANEXO B	FUNÇÃO ESCORE PARA O MODELO DE REGRESSÃO	93
ANEXO C	MATRIZ DE INFORMAÇÃO OBSERVADA	95

INTRODUÇÃO

Em diversas aplicações das áreas do conhecimento humano, tem-se o interesse de modelar características que assumem valores no intervalo unitário $(0, 1)$, como proporções, taxas ou frações. Por exemplo, a porcentagem da renda gasta em alimentação, a proporção de eleitores que votaram em branco por municípios em uma eleição, a taxa de analfabetismo, a porcentagem de itens defeituosos, etc. Um dos modelos mais utilizados para modelar essas características é o modelo beta (KRYSIICKI, 1999). Ferrari e Cribari-Neto (2004) estenderam a família de distribuição beta para um modelo de regressão, assumindo distribuição beta para a variável resposta. Várias extensões do modelo de regressão beta podem ser encontrados na literatura como em Simas, Barreto-Souza e Rocha (2010), Ospina e Ferrari (2010), Ospina e Ferrari (2012), Figueroa-Zúñiga, Arellano-Valle e Ferrari (2013) e Carrasco, Ferrari e Arellano-Valle (2014). Por outro lado, surgiram alguns modelos de regressão alternativo ao modelo de regressão beta. Qiu, Song e Tan (2008) propõem um modelo de regressão baseado na distribuição simplex introduzida por Barndorff-Nielsen e Jørgensen (1991), e recentemente Lemonte e Bazán (2016) propuseram um modelo de regressão baseado em uma classe mais geral que a distribuição S_B de Johnson. Bayes, Bazan e Castro (2017) introduziram o modelo de regressão quantílica paramétrica baseado na distribuição Kumaraswamy (KUMARASWAMY, 1980), e Smithson e Shou (2017) desenvolveram uma família de distribuições para modelar dados no intervalo $(0, 1)$ aplicando a função de distribuição acumulada de uma distribuição sobre a função quantil de outra distribuição.

A proposta deste trabalho é introduzir um novo modelo estatístico para modelar dados no intervalo contínuo $(0, 1)$, baseado em uma transformação de variáveis o qual é formulado da seguinte maneira:

Seja X uma variável aleatória com função distribuição acumulada (f.d.a.) $F_X(x)$ para $x \in \mathbb{R}$. Desde que uma f.d.a. qualquer toma valores no intervalo $(0, 1)$, considere a seguinte

transformação:

$$Y = H\left(\frac{X - \gamma}{\delta}\right), \quad (1.1)$$

em que $\gamma \in \mathbb{R}$, $\delta > 0$ e $H(\cdot)$ é a f.d.a. na forma padrão de uma variável aleatória com o mesmo suporte da variável aleatória X . Nas mesmas condições acima a variável aleatória X tem a seguinte representação estocástica

$$X = \gamma + \delta Q(Y), \quad y \in (0, 1),$$

em que $Q(Y) = H^{-1}(\cdot)$ é a função quantil da f.d.a. $H(\cdot)$. Assim, a função densidade de probabilidade de Y é dada por:

$$f_Y(y; \gamma, \delta) = \delta f_X(\gamma + \delta Q(y)) \left| \frac{dQ(y)}{dy} \right|, \quad y \in (0, 1), \quad (1.2)$$

sendo $f_X(x) = dF_X(x)/dx$, a f.d.p. de X .

Como caso particular desta família de distribuições tem-se quando em (1.1) considera-se a variável aleatória $X \sim N(0, 1)$ com f.d.a. $\Phi(x)$, e $H(\cdot)$ sendo a f.d.a. da distribuição logística padrão ou seja $H(y) = 1/(1 + e^{-y})$, com essas especificações de (1.2) obtém-se a distribuição S_B de Johnson (JOHNSON, 1949), cuja função densidade de probabilidade é dada por:

$$f_Y(y; \gamma, \delta) = \frac{\delta \phi(\gamma + \delta Q(y))}{y(1-y)}, \quad y \in (0, 1), \quad (1.3)$$

em que $\phi(\cdot)$ é a f.d.p. da distribuição normal padrão, e

$$Q(y) = \log\left(\frac{y}{1-y}\right), \quad y \in (0, 1), \quad (1.4)$$

é a função quantil da distribuição logística padrão.

Lemonte e Bazán (2016) introduziram uma ampla classe de distribuições com suporte limitado no intervalo unitário $(0, 1)$, na qual consideraram a transformação em (1.4), e a família de distribuição elíptica (ANDERSON, 1990) com parâmetro de locação igual a *zero* e escala igual a *um* ao invés da normal padrão. A proposta desses autores inclui como caso particular a distribuição S_B de Johnson.

Neste trabalho, é proposto uma nova família de distribuições para modelar dados no intervalo $(0, 1)$ considerando a transformação em (1.1), e a composição da distribuição normal padrão, e a f.d.a. da distribuição valor extremo generalizado (VEG) com parâmetro de locação igual a *zero* e parâmetro de escala igual a *um*. A escolha dessa transformação tem como objetivo obter uma distribuição mais flexível para modelar eventos raros ou extremos ou assimétricos no intervalo $(0, 1)$. Portanto na próxima Seção é apresentada a distribuição valor extremo generalizado.

1.1 Distribuição Valor Extremo Generalizado (VEG)

A distribuição valor extremo generalizado, introduzida por [Jenkinson \(1955\)](#), é uma família de distribuição contínua que foi construída sob a teoria do valor extremo, que é baseada no estudo da distribuição limite das estatísticas do máximo (ou mínimo) normalizado. Essa distribuição tem a capacidade de modelar eventos extremos ou eventos raros, e por isso ela tem sido muito aplicada na modelagem de eventos climáticos extremos, hidrologia, fenômenos meteorológicos, eventos pouco comuns no mercado financeiro, etc.

Definição 1. Uma variável aleatória W tem distribuição valor extremo generalizado com parâmetro de locação *zero*, parâmetro de escala *um* e parâmetro de forma $\lambda \in \mathbb{R}$ se sua f.d.p. é dada por:

$$h(w; \lambda) = \begin{cases} (1 + \lambda w)^{(-\frac{1}{\lambda})-1} \exp\left\{- (1 + \lambda w)^{-\frac{1}{\lambda}}\right\}, & \lambda \neq 0 \\ \exp(-w) \exp[-\exp(-w)], & \lambda = 0. \end{cases} \quad (1.5)$$

A f.d.a. de W é dada por:

$$H(w; \lambda) = \begin{cases} \exp\left\{- (1 + \lambda w)^{-1/\lambda}\right\}, & \lambda \neq 0 \\ \exp[-\exp(-w)], & \lambda = 0, \end{cases} \quad (1.6)$$

em que $w \in [-1/\lambda, +\infty)$ quando $\lambda > 0$, $w \in (-\infty, -1/\lambda]$ quando $\lambda < 0$ e $w \in (-\infty, +\infty)$ quando $\lambda = 0$.

De (1.6), temos que a função quantil de W é dada por:

$$Q(p; \lambda) = \begin{cases} \frac{[-\log(p)]^{-\lambda}-1}{\lambda}, & \lambda \neq 0 \\ -\log[-\log(p)], & \lambda = 0, \end{cases} \quad (1.7)$$

em que $p \in (0, 1)$. Como casos particulares da família de distribuição valor extremo generalizado tem-se, a distribuição Gumbel quando $\lambda = 0$, a distribuição Weibull quando $\lambda < 0$ e a distribuição Frechet quando $\lambda > 0$.

A Figura 1 mostra o comportamento da f.d.p. da distribuição VEG para alguns valores de λ . Assim quando $\lambda > 0$ a distribuição é assimétrica à direita, quando $\lambda = 0$ a distribuição se torna simétrica, e quando $\lambda < 0$ a distribuição é assimétrica à esquerda. A Figura 2 mostra o comportamento da f.d.a. da distribuição VEG com os mesmos valores dos parâmetros que da f.d.p. da distribuição VEG. Note que as probabilidades de um evento tende mais rapidamente para *zero* do que para *um*, enfatizando que tende muito mais rápido quando $\lambda > 0$ do que quando $\lambda = 0$ ou $\lambda < 0$, ou seja probabilidades de eventos menores que 0 se torna zero rapidamente para esses valores dos parâmetros, e ainda com maior rapidez quando λ toma valores positivos. E na Figura 3 é mostrada a f.d.a. da distribuição VEG conjuntamente com a f.d.a. da distribuição logística padrão, note que a distribuição logística é simétrica e a distribuição VEG é assimétrica.

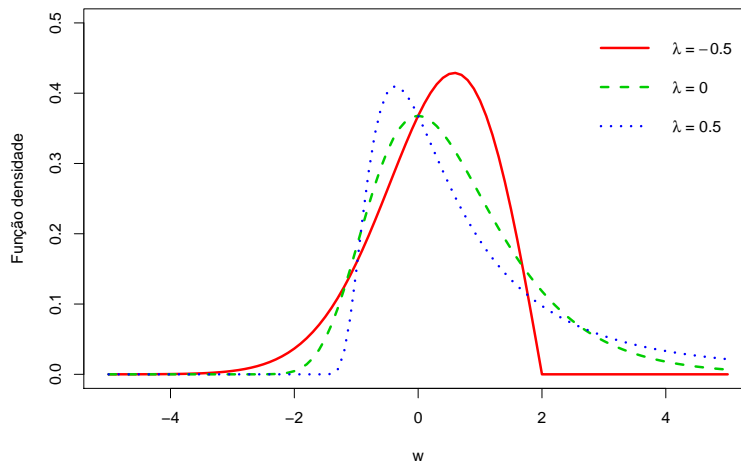


Figura 1 – Função de densidade de probabilidade da distribuição VEG

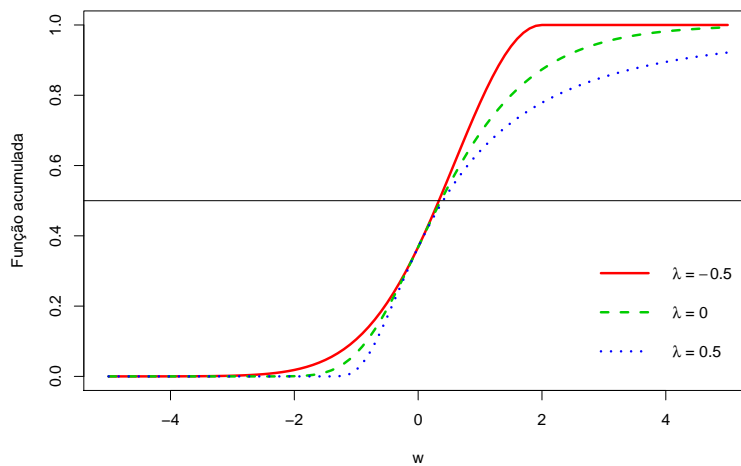


Figura 2 – Função de distribuição acumulada da distribuição VEG

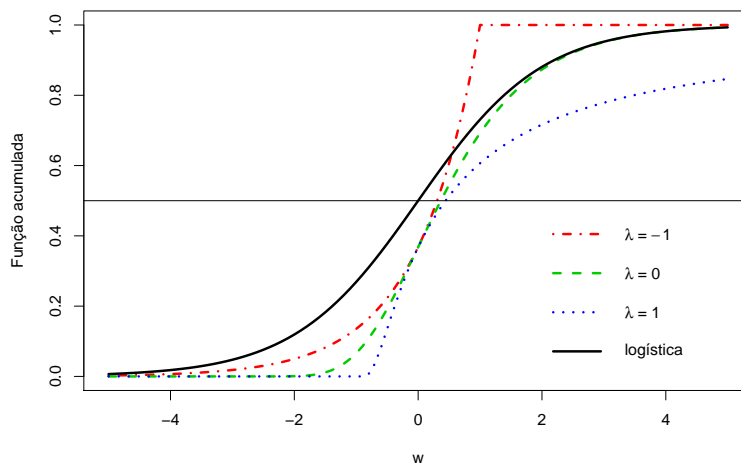


Figura 3 – Função de distribuição acumulada da distribuição VEG e da distribuição logística padrão

1.2 Objetivos

Os objetivos específicos deste trabalho são:

- Propor uma nova família de distribuições para modelar dados no intervalo unitário $(0, 1)$;
- Estudar as propriedades estruturais da nova família;
- Desenvolver procedimentos inferenciais sob a perspectiva clássica e Bayesiana para o novo modelo;
- Estender o modelo proposto para modelos de regressão e desenvolver procedimentos inferenciais na perspectiva clássica e Bayesiana;
- Fazer um estudo de simulação para estudar as propriedades frequentistas dos estimadores.

1.3 Organização do trabalho

O trabalho está organizado da seguinte forma: No Capítulo 2 é proposto o novo modelo e são estudadas suas propriedades. Para o modelo proposto é desenvolvido procedimentos inferenciais baseados no método de máxima verossimilhança. Além disso é conduzido um estudo de simulação para avaliar as propriedades assintóticas dos estimadores dos parâmetros do modelo. Finalmente é desenvolvida uma aplicação a dados reais do modelo. No Capítulo 3 o novo modelo é estendido para um modelo de regressão, onde é estruturado o modelo de regressão e desenvolvidos os procedimentos inferenciais pelo método de máxima verossimilhança. Depois é conduzido um estudo de simulação para avaliar o desempenho dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão homocedástico e heterocedástico. E por último é apresentada uma aplicação a um conjunto de dados de câncer colorretal. No Capítulo 4 é desenvolvido o modelo de regressão sob o enfoque bayesiano. Inferências do modelo são feitas sob este enfoque. São considerados critérios bayesianos para a seleção de modelos, assim como também são estudados métodos para o diagnóstico de influência do modelo. Também é considerado um estudo de simulação para avaliar as propriedades dos estimadores bayesianos dos parâmetros do modelo de regressão, e uma aplicação com os dados de câncer colorretal é considerada. Finalmente no Capítulo 5 são discutidas as considerações deste trabalho, assim como propostas de trabalhos futuros a serem desenvolvidos.

MODELO NORMAL-VEG

Neste Capítulo é introduzido um novo modelo estatístico para modelar dados no intervalo $(0, 1)$. O modelo proposto é uma composição da distribuição normal padrão, e o quantil da distribuição VEG padrão. São estudadas algumas propriedades do modelo e são desenvolvidos procedimentos inferências baseadas no método de máxima verossimilhança.

2.1 Modelo Normal-VEG

A nova família de distribuições é obtida ao considerar em (1.1) a variável X com distribuição normal padrão, cuja função densidade de probabilidade é dada por $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ para $x \in \mathbb{R}$, e $H(\cdot)$ a função distribuição acumulada da distribuição VEG com parâmetro de locação igual a *zero*, parâmetro de escala igual a *um* e parâmetro de forma $\lambda \in \mathbb{R}$. sob essas condições a variável aleatória X em (1.1) tem a seguinte representação estocástica:

$$X = \gamma + \delta Q(Y; \lambda), \quad y \in (0, 1),$$

com

$$Q(y; \lambda) = \begin{cases} \frac{[-\log(y)]^{-\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ -\log[-\log(y)], & \lambda = 0, \end{cases} \quad (2.1)$$

em que $Q(y; \lambda)$ é a função quantil da distribuição VEG para $0 < y < 1$.

De (1.2) tem-se que a f.d.p. da nova família de distribuições é dada por:

$$f_Y(y; \gamma, \delta, \lambda) = \frac{\delta \phi(\gamma + \delta Q(y; \lambda))}{y[-\log(y)]^{\lambda+1}}, \quad y \in (0, 1), \quad (2.2)$$

com $\lambda, \gamma \in \mathbb{R}$, $\delta > 0$ e $Q(y; \lambda)$ é dado como em (2.1). Assim neste estudo denominaremos a essa família de distribuições com f.d.p. em (2.2) de família de distribuições Normal-VEG, e o denotaremos por $Y \sim \text{Normal-VEG}(\gamma, \delta, \lambda)$.

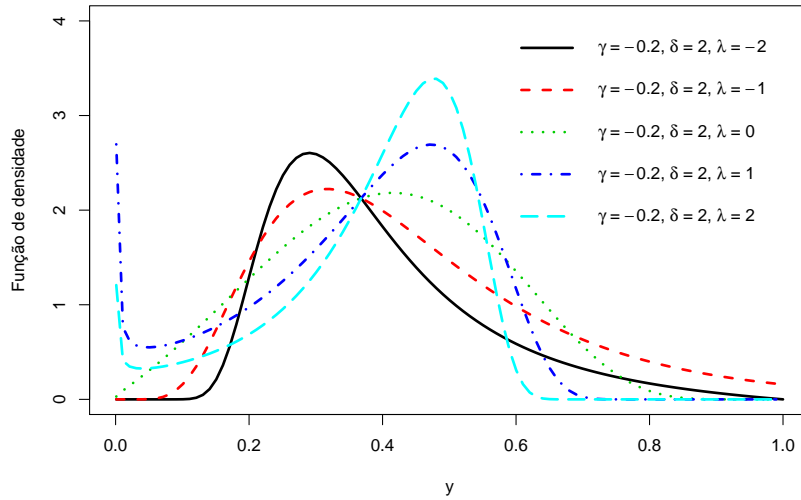


Figura 4 – Gráfico da função densidade de probabilidade da distribuição Normal-VEG com λ variando.

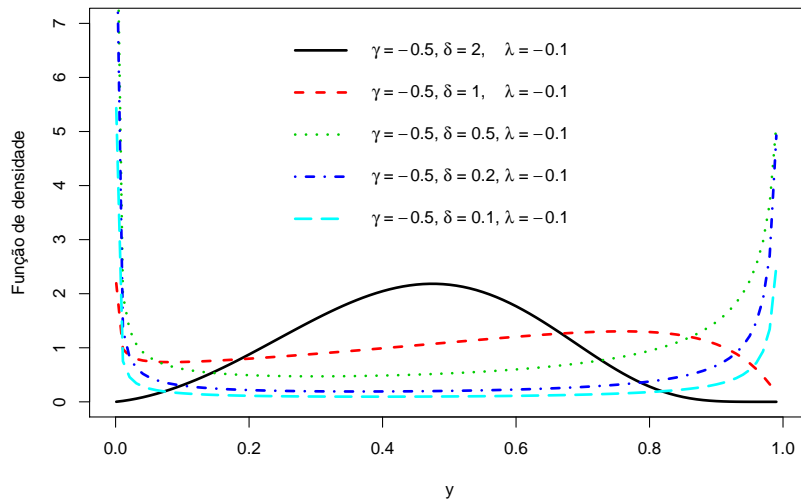


Figura 5 – Gráfico da função densidade de probabilidade da distribuição Normal-VEG com δ variando.

As Figuras 4, 5 e 6 mostram o comportamento da f.d.p. do modelo Normal-VEG para diferentes valores dos parâmetros, mostrando a flexibilidade da distribuição.

A f.d.a. da família Normal-VEG é dada por:

$$F_Y(y) = \Phi(\gamma + \delta Q(y; \lambda)); y \in (0, 1), \quad (2.3)$$

em que $\Phi(\cdot)$ é a f.d.a. da distribuição normal padrão, $\gamma \in \mathbb{R}$, $\delta > 0$ e $Q(y; \lambda)$ é dado em (2.1).

De (2.3), pode-se mostrar que o q -ésimo quantil da distribuição Normal-VEG é dado

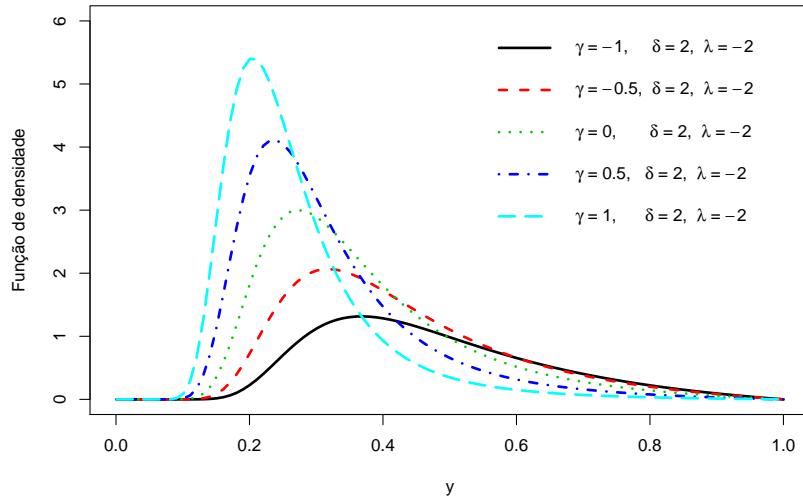


Figura 6 – Gráfico da função densidade de probabilidade da distribuição Normal-VEG, com γ variando.

por:

$$y_q = \begin{cases} \exp \left\{ - \left[1 + \frac{\lambda}{\delta} (x_q - \gamma) \right]^{-\frac{1}{\lambda}} \right\}, & \lambda \neq 0 \\ \exp \left\{ - \exp \left[- \left(\frac{x_q - \gamma}{\delta} \right) \right] \right\}, & \lambda = 0, \end{cases} \quad (2.4)$$

em que x_q é o quantil da distribuição normal padrão e $0 < q < 1$.

De (2.4), pode-se obter a mediana da distribuição Normal-VEG a qual tem a forma:

$$v = \begin{cases} \exp \left\{ - \left[1 + \frac{\lambda}{\delta} (-\gamma) \right]^{-\frac{1}{\lambda}} \right\}, & \lambda \neq 0 \\ \exp \left\{ - \exp \left(\frac{-\gamma}{\delta} \right) \right\}, & \lambda = 0. \end{cases} \quad (2.5)$$

O r -ésimo momento da família de distribuições Normal-VEG, é dado por:

$$E(Y^r) = \int_0^1 y^r \frac{\delta \phi(\gamma + \delta Q(y; \lambda))}{y [-\log(y)]^{\lambda+1}} dy,$$

para $r = 1, 2, \dots$, para o modelo não se tem forma fechada dos momentos, mas pode-se garantir a existência dele devido ao fato do suporte da distribuição ser limitado no intervalo $(0, 1)$.

Na Tabela 1 são apresentados a média e variâncias da distribuição Normal-VEG para diferentes combinações de seus parâmetros, essas medidas foram obtidas numericamente usando a função *integrate* do software R (R Core Team, 2015); observa-se que a primeira coluna de parâmetros é variado o parâmetro λ e mantidos os demais parâmetros constantes, e os valores respectivos das médias e variâncias não variam muito. Agora na segunda coluna de parâmetros é variado o parâmetro γ mantendo os demais parâmetros constantes, esta variação do parâmetro γ influencia na variação da média consideravelmente. E na terceira coluna de parâmetros é variado o parâmetro δ e mantidas os demais parâmetros constantes, observa-se que a variação do parâmetro δ , influencia principalmente na valor da variância. Desta forma, pode-se dizer que

Tabela 1 – Média e variância da Distribuição Normal-VEG obtidas numericamente para diferentes valores dos parâmetros.

Parâmetro	Média	Variância	Parâmetro	Média	Variância	Parâmetro	Média	Variância
$\lambda = 0,5$			$\lambda = 0,1$			$\lambda = 0,5$		
$\gamma = 0,2$	0,3212	0,0279	$\gamma = 2$	0,0921	0,0100	$\gamma = 0,2$	0,3047	0,0622
$\delta = 2$			$\delta = 2$			$\delta = 1$		
$\lambda = 0,1$			$\lambda = 0,1$			$\lambda = 0,5$		
$\gamma = 0,2$	0,3341	0,0266	$\gamma = 1$	0,2105	0,0213	$\gamma = 0,2$	0,3212	0,0279
$\delta = 2$			$\delta = 2$			$\delta = 2$		
$\lambda = 0$			$\lambda = 0,1$			$\lambda = 0,5$		
$\gamma = 0,2$	0,3373	0,0265	$\gamma = 0,5$	0,2856	0,0254	$\gamma = 0,2$	0,3359	0,0144
$\delta = 2$			$\delta = 2$			$\delta = 3$		
$\lambda = -0,2$			$\lambda = 0,1$			$\lambda = 0,5$		
$\gamma = 0,2$	0,3437	0,0267	$\gamma = 0,2$	0,3341	0,0266	$\gamma = 0,2$	0,3532	0,0038
$\delta = 2$			$\delta = 2$			$\delta = 6$		
$\lambda = -0,5$			$\lambda = 0,1$			$\lambda = 0,5$		
$\gamma = 0,2$	0,3538	0,0281	$\gamma = 0$	0,3666	0,0269	$\gamma = 0,2$	0,3573	0,0021
$\delta = 2$			$\delta = 2$			$\delta = 8$		

o parâmetro γ é o parâmetro que está relacionado com a locação do modelo, e o parâmetro δ está relacionado com a dispersão do modelo.

De (2.4) podemos obter uma medida da assimetria do modelo usando o coeficiente quartílico de assimetria ou formula de Bowley (KENNEY; KEEPING, 1954), que é definido como

$$C = \frac{y_{0,75} + y_{0,25} - 2y_{0,5}}{y_{0,75} - y_{0,25}}.$$

O coeficiente quartílico (C) esta sempre compreendido no intervalo $[-1, +1]$. Quando $C = 0$ a distribuição é simétrica, quando $0 < C < 1$ a distribuição é assimétrica à direita, e quando $-1 < C < 0$ a distribuição é assimétrica à esquerda.

A Figura 7 mostra o comportamento do coeficiente quartílico de assimetria para o modelo Normal-VEG para alguns parâmetros. Pode-se observar que para valores de $\lambda < 0$ o modelo é assimétrico a direita, para valores de $\lambda > 0$ o modelo é assimétrico à esquerda, mas também podemos notar que quando o parâmetro $\lambda \rightarrow 0$ e $\gamma \rightarrow 0$ o modelo se torna simétrico. Portanto os parâmetros λ e γ controlam a assimetria da distribuição conjuntamente.

2.2 Inferência

Nesta seção são apresentados os procedimentos inferenciais para os parâmetros do modelo Normal-VEG pelo método de máxima verossimilhança.

Seja uma amostra aleatória Y_1, \dots, Y_n em que $Y_i \sim \text{Normal-VEG}(\gamma, \delta, \lambda)$ para $i = 1, \dots, n$, e com função densidade de probabilidade dada em (2.2). Seja $\boldsymbol{\vartheta} = (\gamma, \delta, \lambda)^T$ o vetor de parâmetros desconhecidos da distribuição, a função log-verossimilhança de $\boldsymbol{\vartheta}$ é dada por

$$\ell(\boldsymbol{\vartheta}) = \sum_{i=1}^n \{ \log \delta + \log \{ \phi(\gamma + \delta Q(y_i; \lambda)) \} - \log(y_i) - (\lambda + 1) \log[-\log(y_i)] \} \quad (2.6)$$

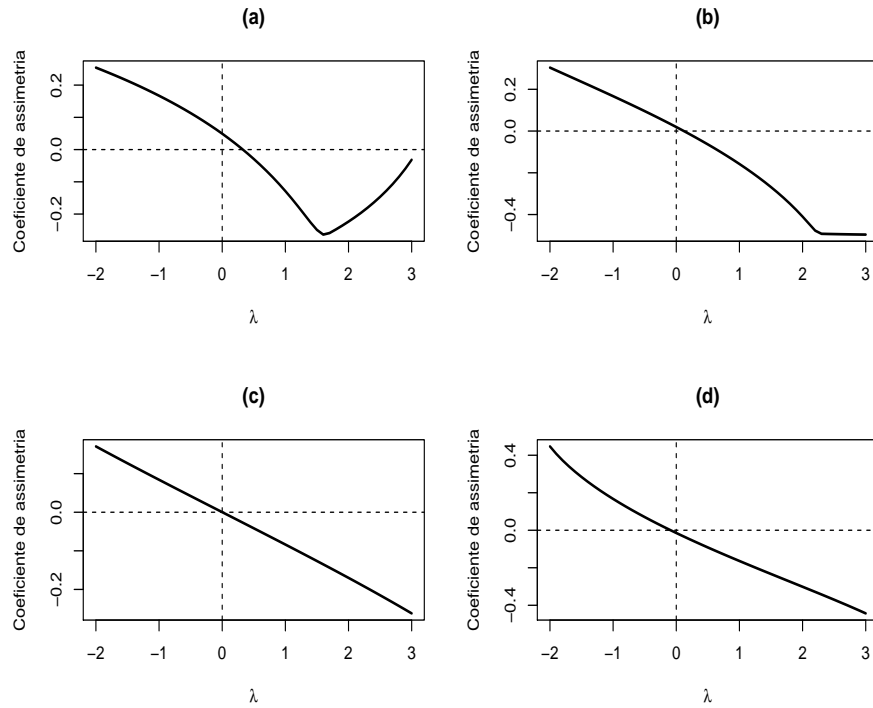


Figura 7 – Coeficiente de assimetria da distribuição Normal-VEG com parâmetros: (a) $\delta = 2$, $\gamma = 0,5$, (b) $\delta = 2$, $\gamma = 0,2$, (c) $\delta = 4$, $\gamma = 0$, (d) $\delta = 2$, $\gamma = -0,2$

em que $Q(y_i; \lambda)$ é a função quantil da distribuição VEG dada em (2.1) para $\lambda \neq 0$.

As primeiras derivadas da função log-verossimilhança $\ell(\boldsymbol{\vartheta})$ em relação a cada um dos parâmetros são dadas por:

$$\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \gamma} = -n\gamma - \delta \sum_{i=1}^n \frac{[-\log(y_i)]^{-\lambda} - 1}{\lambda}, \quad (2.7)$$

$$\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \delta} = \frac{n}{\delta} - \gamma \sum_{i=1}^n \frac{[-\log(y_i)]^{-\lambda} - 1}{\lambda} - \delta \sum_{i=1}^n \left\{ \frac{[-\log(y_i)]^{-\lambda} - 1}{\lambda} \right\}^2, \quad (2.8)$$

e

$$\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \lambda} = \sum_{i=1}^n -\delta[\gamma + \delta Q(y_i; \lambda)] \left\{ -\frac{1}{\lambda^2} \{[-\log(y_i)]^{-\lambda} - 1\} - \frac{1}{\lambda} \log[-\log(y_i)][-\log(y_i)]^{-\lambda} \right\} - \sum_{i=1}^n \log[-\log(y_i)]. \quad (2.9)$$

As estimativas de máxima verossimilhança (EMVs) de $\boldsymbol{\vartheta}$ são os valores que maximizam a função log-verossimilhança dada em (2.6), o que é equivalente a resolver o sistema de equações não lineares

$$\mathbf{U}(\boldsymbol{\vartheta}) = \frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \Big|_{\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}} = \mathbf{0}.$$

Assim, resolvendo a equação $\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \delta} |_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} = 0$, obtém-se uma expressão para $\hat{\delta}$ dada por

$$\hat{\delta} = \sqrt{n \left\{ \sum_{i=1}^n \left\{ \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}} \right\}^2 - \frac{\left\{ \sum_{i=1}^n \frac{[-\log(y_i)]^{-\hat{\lambda}-1}}{\hat{\lambda}} \right\}^2}{n} \right\}^{-1}}. \quad (2.10)$$

Da mesma forma resolvendo $\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \gamma} |_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} = 0$, tem-se que

$$\hat{\gamma} = -\frac{\hat{\delta}}{n} \sum_{i=1}^n \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}}, \quad (2.11)$$

em que $\hat{\delta}$ é dado em (2.10). Note que quando $\hat{\lambda} \rightarrow 0$ tem-se que:

$$\hat{\gamma} = -\frac{\hat{\delta}}{n} \sum_{i=1}^n -\log[-\log(y_i)], \quad (2.12)$$

e também

$$\hat{\delta} = \sqrt{n \left\{ \sum_{i=1}^n \{-\log[-\log(y_i)]\}^2 - \frac{\{\sum_{i=1}^n -\log[-\log(y_i)]\}^2}{n} \right\}^{-1}}. \quad (2.13)$$

Por último substituindo (2.10) e (2.11) em (2.9) e resolvendo a equação $\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \lambda} |_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} = 0$ em $\hat{\lambda}$ obtém-se

$$-\sum_{i=1}^n \hat{\delta} [\hat{\gamma} + \hat{\delta} Q(y_i; \hat{\lambda})] \left\{ -\frac{1}{\hat{\lambda}^2} \{[-\log(y_i)]^{-\hat{\lambda}} - 1\} - \frac{1}{\hat{\lambda}} \log[-\log(y_i)] [-\log(y_i)]^{-\hat{\lambda}} \right\} = \sum_{i=1}^n \log[-\log(y_i)]. \quad (2.14)$$

Como não é possível encontrar analiticamente uma expressão fechada para $\hat{\lambda}$ pode-se utilizar algum método numérico para resolver a equação acima. Um dos métodos mais usados é o método de Newton-Raphson. Esse método aparece como subrotina em alguns softwares como no aplicativo R.

Inferência sobre os parâmetros são feitas usualmente por médio da normalidade assintótica dos EMV (HINKLEY; COX, 1979). Assim para amostras suficientemente grandes e sob certas condições de regularidade, pode-se mostrar que

$$\hat{\boldsymbol{\vartheta}} \sim N_3(\boldsymbol{\vartheta}, \mathbf{I}(\boldsymbol{\vartheta})^{-1}) \quad (2.15)$$

em que a matriz de informação de Fisher $\mathbf{I}(\boldsymbol{\vartheta})$, é dada por

$$\mathbf{I}(\boldsymbol{\vartheta}) = E \left[-\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} \right].$$

A expressão $\mathbf{I}(\boldsymbol{\vartheta})$ para o modelo Normal-VEG não tem forma fechada. Mas, ela pode ser aproximada pela matriz de informação observada $\mathbf{J}(\boldsymbol{\vartheta}) = -\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T}$ avaliada em $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}$. Assim,

$$\mathbf{J}(\hat{\boldsymbol{\vartheta}}) = \begin{bmatrix} -\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \gamma^2} & -\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \gamma \partial \delta} & -\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \gamma \partial \lambda} \\ -\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \delta^2} & -\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \delta \partial \lambda} & -\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \lambda^2} \end{bmatrix}_{\hat{\boldsymbol{\vartheta}}},$$

em que os elementos da matriz de informação estão apresentados no Anexo A.

Considerando a normalidade assintótica dos EMV de $\boldsymbol{\vartheta}$, o intervalo assintótico de $100(1 - \alpha)\%$ de confiança para esses parâmetros é dado por:

$$(\hat{\vartheta}_j - z_{1-\alpha/2} EP(\hat{\vartheta}_j); \hat{\vartheta}_j + z_{1-\alpha/2} EP(\hat{\vartheta}_j)), j = 1, 2, 3,$$

em que $z_{1-\alpha/2}$ é o quantil $1 - \alpha/2$ da distribuição normal padrão e $EP(\hat{\vartheta}_j)$ é o erro padrão do EMV de ϑ_j obtida pela raiz quadrada do j -ésimo elemento da diagonal da inversa da matriz de informação observada.

2.3 Estudo de simulação

Um estudo de simulação foi realizado para avaliar as propriedades assintóticas dos estimadores de máxima verossimilhança (EMV) dos parâmetros do modelo Normal-VEG. Para isto foram geradas amostras de tamanhos $n = 50, 100, 200, 400$ e 800 do modelo, com as seguintes configurações de parâmetros:

- $\gamma = 2$ e $\delta = 10$ com $\lambda = -0,2; 0$ e $0,2$;
- $\gamma = 0,1$ e $\delta = 2$ com $\lambda = -0,1; 0$ e $0,1$;
- $\gamma = -0,2$ e $\delta = 6$ com $\lambda = -0,5; 0$ e $0,5$; e
- $\gamma = -0,1$ e $\delta = 1$ com $\lambda = -0,2; 0$ e $0,2$.

As amostras foram geradas mediante a seguinte expressão:

$$y_i = \begin{cases} \exp \left\{ -\left[1 + \frac{\lambda}{\delta} (x_i - \gamma) \right]^{-\frac{1}{\lambda}} \right\}, & \lambda \neq 0 \\ \exp \left\{ -\exp \left[-\left(\frac{x_i - \gamma}{\delta} \right) \right] \right\}, & \lambda = 0, \end{cases}$$

em que x_i foi gerada da distribuição $N(0, 1)$, para $i = 1, \dots, n$. Para cada configuração de parâmetros e tamanho de amostra foram simulados 1000 amostras, e para cada amostra foi obtida a estimativa de máxima verossimilhança dos parâmetros do modelo, o erro padrão (EP) e o intervalo de 95% de confiança respectivo. Das EMVs foram obtidos as médias do erro padrão (MEP), os desvios padrões (DP), os vies, a raiz quadrada do erro quadrático médio (REQM) e a probabilidade de cobertura (PC) para cada um dos parâmetros do modelo. Essas quantidades foram calculadas da seguinte forma:

$$\text{Média} = \frac{1}{1000} \sum_{i=1}^{1000} \hat{\vartheta}_j^{(i)}, \quad \text{MEP} = \frac{1}{1000} \sum_{i=1}^{1000} EP_j^{(i)},$$

$$\text{Viés}(\hat{\vartheta}_j) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\vartheta}_j^{(i)} - \vartheta_j) \text{ e } \text{REQM}(\hat{\vartheta}_j) = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\vartheta}_j^{(i)} - \vartheta_j)^2}$$

em que EP_j é o erro padrão correspondente à estimativa do parâmetro j , que é obtida calculando a raiz quadrada do j -ésimo elemento da diagonal da inversa da matriz de informação observada, para $j = 1, 2, 3$ e $\hat{\vartheta}_j^{(i)}$ é a estimativa de ϑ_j na i -ésima amostra para $\vartheta_j \in \boldsymbol{\vartheta} = (\lambda, \gamma, \delta)^\top$. Os resultados são apresentados nas tabelas seguintes:

- Na Tabela 2 são apresentados os resultados da simulação para o modelo Normal-VEG com parâmetros $\gamma = 2$, $\delta = 10$, e $\lambda = -0, 2; 0; 0, 2$. Observa-se para os três casos que à medida que o tamanho da amostra aumenta, as EMVs são próximas aos valores verdadeiros dos parâmetros em média. Tem-se também o viés e a raiz do erro quadrático médio tendem para zero. Além disso, a média dos erros padrões e os desvios padrões para cada estimativa são aproximadamente iguais à medida que o tamanho da amostra aumenta, e o valor da probabilidade de cobertura se aproxima do valor de 95% como era esperado. Esses resultados indicam que as propriedades assintóticas dos EMV dos parâmetros do modelo estão satisfeitas.
- Na Tabela 3 são apresentados os resultados da simulação do modelo Normal-VEG com parâmetros $\gamma = 0, 1$, $\delta = 2$, e $\lambda = -0, 1; 0; 0, 1$. Os resultados em termos gerais são os mesmos do item anterior, mas pode-se notar aqui que já com uma amostra de tamanho 50 o viés fica pequeno e com o aumento da amostra se torna zero para os três casos do parâmetro λ , fornecendo com isso boas estimativas para os parâmetros do modelo para estes valores dos parâmetros.
- Na Tabela 4 são apresentados os resultados da simulação do modelo Normal-VEG com parâmetros $\gamma = -0, 2$, $\delta = 6$, e $\lambda = -0, 5; 0; 0, 5$. Para esta combinação de parâmetros também tem-se resultados similares ao primeiro item, ou seja os EMV dos parâmetros possuem boas propriedades assintóticas.
- Na Tabela 5 são apresentados os resultados da simulação do modelo Normal-VEG com parâmetros $\gamma = -0, 1$, $\delta = 1$, e $\lambda = -0, 2; 0; 0, 2$. Neste caso consideramos os mesmos valores para λ como no primeiro item, mas diminuindo os valores para os parâmetros γ e δ ; pode-se notar que para essa combinação de valores dos parâmetros tem-se melhores resultados já a partir de uma amostra de tamanho 50 conforme mostra a primeira tabela de resultados. Isto pode-se notar melhor quando $\lambda = 0$.

Em conclusão, é mostrado que os estimadores de máxima verossimilhança dos parâmetros do modelo proposto possuem boas propriedades assintóticas. Notando também que para os quatro casos, para um tamanho de amostra 200 obtêm-se PC de 95%, que é o esperado.

Tabela 2 – Resultados da simulação do modelo Normal-VEG com parâmetros $\lambda = -0, 2; 0; 0, 2$, com $\gamma = 2$ e $\delta = 10$.

n	Par.	$\lambda = -0, 2$						$\lambda = 0$						$\lambda = 0, 2$					
		Média	MEP	DP	Viés	REQM	PC	Média	MEP	DP	Viés	REQM	PC	Média	MEP	DP	Viés	REQM	PC
50	λ	-0,168	1,329	1,253	0,032	1,253	0,963	0,023	1,278	1,202	0,023	1,201	0,967	0,209	1,226	1,153	0,009	1,152	0,970
	γ	2,095	0,335	0,345	0,095	0,358	0,976	2,094	0,334	0,344	0,094	0,357	0,976	2,091	0,332	0,342	0,091	0,353	0,975
	δ	10,735	3,029	3,072	0,735	3,157	0,959	10,707	2,980	3,016	0,707	3,097	0,956	10,666	2,926	2,953	0,666	3,026	0,955
100	λ	-0,163	0,896	0,868	0,037	0,868	0,966	0,029	0,862	0,835	0,029	0,835	0,966	0,221	0,827	0,802	0,021	0,801	0,965
	γ	2,049	0,224	0,231	0,049	0,236	0,962	2,048	0,224	0,230	0,048	0,235	0,962	2,047	0,223	0,229	0,047	0,233	0,962
	δ	10,403	1,974	2,012	0,403	2,051	0,962	10,383	1,945	1,981	0,383	2,017	0,959	10,363	1,915	1,949	0,363	1,981	0,959
200	λ	-0,178	0,616	0,605	0,022	0,605	0,950	0,018	0,592	0,582	0,018	0,582	0,950	0,214	0,568	0,558	0,014	0,558	0,950
	γ	2,018	0,154	0,156	0,018	0,157	0,949	2,018	0,153	0,156	0,018	0,157	0,949	2,017	0,153	0,156	0,017	0,156	0,949
	δ	10,185	1,328	1,370	0,185	1,382	0,950	10,176	1,310	1,351	0,176	1,362	0,952	10,165	1,291	1,330	0,165	1,340	0,952
400	λ	-0,208	0,430	0,428	-0,008	0,428	0,948	-0,010	0,413	0,412	-0,010	0,412	0,948	0,189	0,396	0,395	-0,011	0,395	0,948
	γ	2,006	0,107	0,107	0,006	0,107	0,955	2,005	0,107	0,107	0,005	0,107	0,954	2,005	0,107	0,107	0,005	0,107	0,955
	δ	10,050	0,915	0,938	0,050	0,938	0,946	10,046	0,904	0,925	0,046	0,926	0,946	10,042	0,891	0,911	0,042	0,912	0,947
800	λ	-0,196	0,302	0,300	0,004	0,300	0,949	0,003	0,290	0,289	0,003	0,289	0,949	0,202	0,279	0,277	0,002	0,277	0,951
	γ	2,005	0,076	0,075	0,005	0,075	0,954	2,005	0,075	0,075	0,005	0,075	0,953	2,005	0,075	0,075	0,005	0,075	0,953
	δ	10,041	0,643	0,650	0,041	0,651	0,943	10,039	0,635	0,642	0,039	0,642	0,943	10,036	0,627	0,633	0,036	0,633	0,943

Tabela 3 – Resultados da simulação do modelo Normal-VEG com $\lambda = -0, 1; 0; 0, 1$, com $\gamma = 0, 1$ e $\delta = 2$.

n	Par.	$\lambda = -0, 1$						$\lambda = 0$						$\lambda = 0, 1$					
		Média	MEP	DP	Viés	REQM	PC	Média	MEP	DP	Viés	REQM	PC	Média	MEP	DP	Viés	REQM	PC
50	λ	-0,090	0,256	0,243	0,010	0,243	0,960	0,004	0,256	0,242	0,004	0,242	0,965	0,097	0,254	0,240	-0,003	0,240	0,968
	γ	0,103	0,155	0,151	0,003	0,151	0,954	0,105	0,155	0,151	0,005	0,151	0,954	0,106	0,155	0,151	0,006	0,151	0,955
	δ	2,068	0,211	0,215	0,068	0,226	0,956	2,067	0,212	0,216	0,067	0,226	0,957	2,067	0,212	0,217	0,067	0,227	0,957
100	λ	-0,091	0,173	0,167	0,009	0,167	0,967	0,006	0,172	0,167	0,006	0,167	0,966	0,103	0,171	0,166	0,003	0,166	0,963
	γ	0,099	0,109	0,105	-0,001	0,105	0,957	0,099	0,109	0,105	-0,001	0,105	0,957	0,100	0,109	0,105	0,000	0,105	0,958
	δ	2,035	0,145	0,147	0,035	0,151	0,952	2,034	0,146	0,148	0,034	0,151	0,951	2,034	0,146	0,148	0,034	0,152	0,952
200	λ	-0,095	0,119	0,117	0,005	0,117	0,952	0,004	0,118	0,116	0,004	0,116	0,950	0,102	0,118	0,115	0,002	0,115	0,950
	γ	0,096	0,077	0,076	-0,004	0,076	0,954	0,097	0,077	0,076	-0,003	0,076	0,954	0,097	0,077	0,076	-0,003	0,076	0,954
	δ	2,013	0,101	0,102	0,013	0,103	0,945	2,013	0,102	0,103	0,013	0,104	0,943	2,013	0,102	0,103	0,013	0,104	0,940
400	λ	-0,101	0,083	0,083	-0,001	0,083	0,946	0,002	0,083	0,082	-0,002	0,082	0,948	0,097	0,082	0,082	-0,003	0,082	0,946
	γ	0,099	0,054	0,054	-0,001	0,054	0,955	0,099	0,054	0,054	-0,001	0,054	0,955	0,099	0,054	0,053	-0,001	0,053	0,954
	δ	2,006	0,071	0,069	0,006	0,070	0,956	2,006	0,071	0,070	0,006	0,070	0,959	2,006	0,072	0,070	0,006	0,070	0,958
800	λ	-0,099	0,058	0,058	0,001	0,058	0,948	0,000	0,058	0,058	0,000	0,058	0,949	0,100	0,058	0,057	0,000	0,057	0,953
	γ	0,100	0,038	0,038	0,000	0,038	0,957	0,100	0,038	0,038	0,000	0,038	0,960	0,100	0,038	0,038	0,000	0,038	0,960
	δ	2,003	0,050	0,049	0,003	0,049	0,954	2,003	0,050	0,049	0,003	0,049	0,956	2,003	0,051	0,050	0,003	0,050	0,955

Tabela 4 – Resultados da simulação do modelo Normal-VEG com $\lambda = -0,5$; 0 ; $0,5$, com $\gamma = -0,2$ e $\delta = 6$.

n	Par.	$\lambda = -0,5$						$\lambda = 0$						$\lambda = 0,5$					
		Média	MEP	DP	Viés	REQM	PC	Média	MEP	DP	Viés	REQM	PC	Média	MEP	DP	Viés	REQM	PC
50	λ	-0,457	0,749	0,710	0,043	0,711	0,963	0,011	0,767	0,727	0,011	0,726	0,965	0,478	0,775	0,733	-0,022	0,733	0,968
	γ	-0,208	0,155	0,152	-0,008	0,152	0,953	-0,205	0,155	0,152	-0,005	0,152	0,954	-0,203	0,155	0,152	-0,003	0,152	0,953
	δ	6,192	0,655	0,654	0,192	0,681	0,962	6,199	0,649	0,650	0,199	0,680	0,957	6,204	0,643	0,647	0,204	0,678	0,954
100	λ	-0,465	0,505	0,488	0,035	0,489	0,969	0,018	0,517	0,501	0,018	0,501	0,966	0,501	0,522	0,506	0,001	0,506	0,963
	γ	-0,207	0,109	0,106	-0,007	0,106	0,956	-0,206	0,109	0,106	-0,006	0,106	0,956	-0,204	0,109	0,107	-0,004	0,107	0,958
	δ	6,093	0,451	0,441	0,093	0,450	0,960	6,098	0,447	0,438	0,098	0,448	0,960	6,101	0,443	0,436	0,101	0,447	0,958
200	λ	-0,480	0,347	0,341	0,020	0,341	0,952	0,011	0,355	0,349	0,011	0,349	0,950	0,502	0,359	0,352	0,002	0,351	0,950
	γ	-0,206	0,077	0,076	-0,006	0,076	0,956	-0,205	0,077	0,076	-0,005	0,077	0,955	-0,205	0,077	0,077	-0,005	0,077	0,956
	δ	6,034	0,315	0,307	0,034	0,308	0,953	6,037	0,311	0,304	0,037	0,306	0,952	6,038	0,309	0,302	0,038	0,305	0,950
400	λ	-0,501	0,242	0,241	-0,001	0,241	0,946	-0,006	0,248	0,247	-0,006	0,247	0,948	0,490	0,250	0,249	-0,010	0,249	0,948
	γ	-0,202	0,054	0,054	-0,002	0,054	0,954	-0,202	0,054	0,054	-0,002	0,054	0,954	-0,201	0,054	0,054	-0,001	0,054	0,955
	δ	6,018	0,221	0,211	0,018	0,212	0,956	6,019	0,219	0,210	0,019	0,210	0,956	6,019	0,217	0,208	0,019	0,209	0,955
800	λ	-0,496	0,170	0,169	0,004	0,169	0,947	0,001	0,174	0,173	0,001	0,173	0,949	0,499	0,176	0,175	-0,001	0,174	0,953
	γ	-0,201	0,038	0,038	-0,001	0,038	0,955	-0,200	0,038	0,038	0,000	0,038	0,957	-0,200	0,038	0,038	0,000	0,038	0,954
	δ	6,008	0,156	0,150	0,008	0,150	0,956	6,009	0,154	0,149	0,009	0,149	0,955	6,009	0,153	0,147	0,009	0,148	0,956

Tabela 5 – Resultados da simulação do modelo Normal-VEG com $\lambda = -0, 2; 0; 0, 2$, com $\gamma = -0, 1$ e $\delta = 1$.

n	Par.	$\lambda = -0, 2$						$\lambda = 0$						$\lambda = 0, 2$						
		Média	MEP	DP	Viés	REQM	PC	Média	MEP	DP	Viés	REQM	PC	Média	MEP	DP	Viés	REQM	PC	
50	λ	-0,185	0,120	0,114	0,015	0,115	0,959	0,001	0,128	0,122	0,001	0,122	0,964	0,188	0,126	0,118	-0,012	0,118	0,118	0,964
	γ	-0,108	0,154	0,150	-0,008	0,150	0,957	-0,102	0,155	0,151	-0,002	0,151	0,954	-0,096	0,155	0,151	0,004	0,151	0,151	0,950
	δ	1,032	0,108	0,108	0,032	0,113	0,959	1,034	0,106	0,107	0,034	0,112	0,953	1,034	0,105	0,107	0,034	0,107	0,112	0,954
100	λ	-0,190	0,081	0,078	0,010	0,078	0,969	0,003	0,086	0,084	0,003	0,084	0,966	0,195	0,084	0,081	-0,005	0,084	0,082	0,967
	γ	-0,107	0,108	0,105	-0,007	0,105	0,953	-0,104	0,109	0,106	-0,004	0,106	0,955	-0,101	0,109	0,106	-0,001	0,106	0,106	0,955
	δ	1,015	0,074	0,073	0,015	0,074	0,959	1,017	0,073	0,072	0,017	0,074	0,955	1,018	0,072	0,072	0,018	0,072	0,074	0,956
200	λ	-0,195	0,055	0,054	0,005	0,054	0,954	0,002	0,059	0,058	0,002	0,058	0,950	0,197	0,058	0,056	-0,003	0,056	0,056	0,951
	γ	-0,106	0,076	0,076	-0,006	0,076	0,954	-0,105	0,077	0,076	-0,005	0,076	0,956	-0,103	0,077	0,076	-0,003	0,076	0,076	0,957
	δ	1,006	0,052	0,051	0,006	0,051	0,951	1,006	0,051	0,050	0,006	0,051	0,949	1,008	0,051	0,050	0,008	0,051	0,051	0,949
400	λ	-0,201	0,038	0,038	-0,001	0,038	0,947	-0,001	0,041	0,041	-0,001	0,041	0,948	0,197	0,040	0,040	-0,003	0,040	0,040	0,951
	γ	-0,101	0,054	0,054	-0,001	0,054	0,952	-0,101	0,054	0,054	-0,001	0,054	0,955	-0,100	0,054	0,054	0,000	0,054	0,054	0,955
	δ	1,004	0,037	0,035	0,004	0,035	0,956	1,003	0,036	0,034	0,003	0,035	0,957	1,005	0,036	0,034	0,005	0,034	0,034	0,958
800	λ	-0,203	0,027	0,025	-0,003	0,026	0,954	0,000	0,029	0,029	0,000	0,029	0,949	0,200	0,028	0,027	0,000	0,027	0,027	0,955
	γ	-0,099	0,038	0,040	0,001	0,040	0,952	-0,100	0,038	0,038	0,000	0,038	0,958	-0,100	0,038	0,039	0,000	0,039	0,039	0,957
	δ	1,004	0,026	0,025	0,004	0,025	0,957	1,002	0,025	0,024	0,002	0,024	0,957	1,004	0,025	0,024	0,004	0,024	0,024	0,961

2.4 Aplicação

Nesta seção o objetivo é mostrar a aplicabilidade da distribuição em modelar dados limitados no intervalo $(0, 1)$. Para isso foi considerado o conjunto de dados introduzidos por [Siegel, DeSantis e Jemal \(2014\)](#), que trata de dados de câncer colorretal nos Estados Unidos Americanos (EUA) entre os anos 2006 e 2010. Dos dados considerou-se a taxa de incidência de câncer colorretal como a variável de interesse (Y), onde essa taxa de incidência foi calculada pelos autores por 100.000 habitantes, isto é

$$\text{Taxa de incidência} = \frac{\text{número de casos de incidência de câncer colorretal}}{100.000 \text{ habitantes}}$$

Na Tabela 6 são apresentados as medidas descritivas da taxa de incidência de câncer colorretal nos EUA, e pode-se ver que a taxa de incidência tem uma distribuição não simétrica com certa assimetria à direita, isto também é conferido pelo histograma na Figura 8.

Tabela 6 – Medidas descritivas da taxa de incidência de Câncer Colorretal nos EUA

Min.	1 Qu.	Mediana	Média	3 Qu.	Máx.
0,240	0,385	0,460	0,469	0,540	0,766

Como os dados pertencem ao intervalo $(0, 1)$, é assumido que $Y \sim \text{Normal-VEG}(\gamma, \delta, \lambda)$. Os parâmetros foram estimados pelo método de máxima verossimilhança.

De (2.10)-(2.14), obtém-se as estimativas de máxima verossimilhança para os parâmetros do modelo, o qual é mostrado na Tabela 7, conjuntamente com o erro padrão. E com a finalidade de comparar modelos, os dados também foram ajustados com o modelo S_B de Johnson, o modelo GJS-t-Student com 4 graus de liberdade, esse introduzido por [Lemonte e Bazán \(2016\)](#), e com o modelo Beta.

Tabela 7 – Estimativas de máxima verossimilhança e erro padrão dos parâmetros dos modelos Normal-VEG, S_B de Johnson, GJS-t-student(4) e Beta para a taxa de incidência de câncer colorretal.

Parâmetro	Modelo				
	Normal-VEG	S_B de Johnson	GJS-t-Student	Beta	
EMV	γ	-0,898	0,269	0,411	8,95
	δ	3,977	2,125	2,603	10,09
	λ	-0,826			
EP	γ	0,080	0,069	0,069	0,841
	δ	0,281	0,101	0,101	0,952
	λ	0,177			

Pelos critérios de comparação de modelos Informação de Akaike (AIC) ([BOZDOGAN, 1987](#)), e o Critério Bayesiano de Schwarz (BIC) ([SCHWARZ et al., 1978](#)) apresentados na Tabela 8, indicam que o modelo Normal-VEG é o que melhor ajustou esses dados, e isto pode ser observado também na Figura 8.

Tabela 8 – Critérios de comparação de modelos AIC e BIC para os modelos Normal-VEG, S_B de Johnson, GJS-t-Student(4) e Beta para a taxa de incidência de câncer colorretal.

Modelo	Critério	
	AIC	BIC
Normal-VEG	-345,247	-335,067
S_B Johnson	-338,787	-332,000
GJS-t-Student(4)	-331,614	-324,827
Beta	-338.004	-331.216

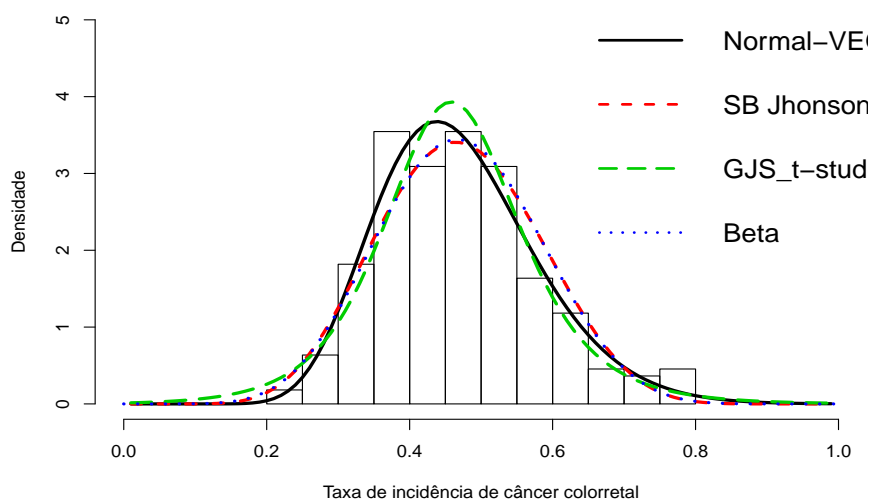


Figura 8 – Ajuste do Modelo Normal-VEG, do Modelo S_B Johnson, do Modelo GJS-t-Student, e do Modelo Beta aos dados

2.5 Conclusão

Neste capítulo foi proposto um novo modelo para modelar dados no intervalo $(0, 1)$, que foi denominado de modelo Normal-VEG. Esse modelo foi baseado em uma transformação de variáveis como na distribuição S_B de Johnson, mas considerando como transformação o quantil da distribuição valor extremo generalizado ao invés do quantil da distribuição logística com o objetivo de obter maior flexibilidade dentre os modelos que modelam este tipo de dados. Para o modelo proposto foram estudadas as suas propriedades, mostrando que os parâmetros λ e γ são os que modelam a assimetria do modelo. A estimação dos parâmetros do modelo foi pelo método de máxima verossimilhança. Foi realizado um estudo de simulação do modelo, na qual foi verificado que os estimadores dos parâmetros do modelo possuem boas propriedades assintoticamente, e também indicando que para amostras de tamanho maior que 200 as propriedades estão satisfeitas. E, por último, um conjunto de dados reais de câncer colorretal foi ajustado com vários modelos notando que o modelo proposto é o que melhor ajusta esse conjunto de dados.

MODELO DE REGRESSÃO NORMAL-VEG

Neste capítulo o modelo normal-VEG é estendido para um modelo de regressão. São desenvolvidos procedimentos inferenciais baseados na teoria de máxima verossimilhança e também são avaliadas as propriedades dos estimadores de máxima verossimilhança por meio de um estudo de simulação. Além disso, uma aplicação do modelo de regressão com dados reais é considerado.

3.1 O modelo Normal-VEG parametrizado na mediana

Geralmente os modelos de regressão são construídos para modelar a média da variável resposta ou uma função dela em função de covariáveis (ou variáveis preditoras) e parâmetros desconhecidos. Neste caso, para a distribuição Normal-VEG não temos uma expressão explícita para a média da variável resposta, mas a mediana (v) é explícita, e é dada em (2.5), razão pela qual é construído um modelo de regressão para modelar a mediana da variável resposta. Para isso, a função densidade de probabilidade em (2.2) é reparametrizada na mediana.

De (2.5), pode-se obter que

$$\gamma = -\delta Q(v; \lambda),$$

com $Q(v; \lambda)$ dada em (2.1), e com essa reparametrização tem-se que a f.d.p. em (2.2) pode ser expressada da seguinte forma

$$f_Y(y) = \frac{\delta \phi(\delta[Q(y; \lambda) - Q(v; \lambda)])}{y[-\log(y)]^{\lambda+1}}, \quad y \in (0, 1), \quad (3.1)$$

em que $0 < v < 1$, $\delta > 0$ e $\lambda \in \mathbb{R}$.

Por outro lado, pode-se provar que δ é o parâmetro de dispersão do modelo em (3.1), e no modelo de regressão é assumido que ele varia de uma observação para outra.

3.1.1 Parâmetro de dispersão

Um conceito utilizado para comparar distribuições de probabilidade quanto a sua dispersão é a de ordenação dispersiva (*dispersive ordering*). Recentemente [Townsend e Colonijs \(2005\)](#) desenvolveram um conceito equivalente chamado de ordem de dispersão quantílica (*Quantil Spread Order*). Assim, para demonstrar que o parâmetro δ é realmente o parâmetro de dispersão, é utilizado este conceito.

Definição 2. Seja Z uma variável aleatória com função distribuição F e função quantil F^{-1} , a ordem de dispersão quantílica de Z é dado por

$$QS_Z(u) = F^{-1}(1-u) - F^{-1}(u), \quad 0 < u < 0,5, \quad (3.2)$$

Por outro lado, seja Z_1 e Z_2 variáveis aleatórias, com funções quantil F_1^{-1} e F_2^{-1} respectivamente; Z_1 é menor que Z_2 em ordem dispersiva quantílica se

$$QS_{Z_1}(u) \leq QS_{Z_2}(u), \quad 0 < u < 0.5, \quad (3.3)$$

e é denotado por $Z_1 \leq_{QS} Z_2$, em que \leq_{QS} implica comparação em variabilidade.

E como em [Lemonte e Bazán \(2016\)](#), tem-se a seguinte proposição, onde é mostrado que δ é o parâmetro de dispersão.

Proposição 1. Seja $Y_1 \sim \text{Normal-VEG}(v, \delta_1, \lambda)$ e $Y_2 \sim \text{Normal-VEG}(v, \delta_2, \lambda)$, então Y_1 é menor que Y_2 em ordem de dispersão quantílica ($Y_1 \leq_{QS} Y_2$), se e só se $\delta_2 \leq \delta_1$.

Demonstração. Se $Y \sim \text{Normal-VEG}(v, \delta, \lambda)$ com função quantil dada em (2.4), então a ordem de dispersão quantílica de Y é dado por

$$QS_Y(u) = \exp \left\{ - \left\{ 1 + \frac{\lambda}{\delta} [x_{1-u} + \delta Q(v; \lambda)] \right\}^{-\frac{1}{\lambda}} \right\} - \exp \left\{ - \left\{ 1 + \frac{\lambda}{\delta} [x_u + \delta Q(v; \lambda)] \right\}^{-\frac{1}{\lambda}} \right\},$$

para $0 < u < 0.5$, em que $Q(\cdot)$ é dado em (2.1), x_u é o u -ésimo quantil de $X \sim N(0, 1)$.

Então, Y_1 é menor que Y_2 em ordem de dispersão quantílica ou $Y_1 \leq_{QS} Y_2$, se

$$QS_{Y_1}(u) \leq QS_{Y_2}(u), \quad 0 < u < 0.5,$$

ou seja tem-se que demonstrar que $QS_{Y_1}(u) - QS_{Y_2}(u) \leq 0$, se e só se $\delta_2 \leq \delta_1$, então tem-se que

$$\begin{aligned} QS_{Y_1}(u) - QS_{Y_2}(u) &= \exp \left\{ - \left[1 + \frac{\lambda}{\delta_1} [x_{1-u} + \delta_1 Q(v; \lambda)] \right]^{-\frac{1}{\lambda}} \right\} - \exp \left\{ - \left[1 + \frac{\lambda}{\delta_1} [x_u + \delta_1 Q(v; \lambda)] \right]^{-\frac{1}{\lambda}} \right\} \\ &\quad - \exp \left\{ - \left[1 + \frac{\lambda}{\delta_2} [x_{1-u} + \delta_2 Q(v; \lambda)] \right]^{-\frac{1}{\lambda}} \right\} + \exp \left\{ - \left[1 + \frac{\lambda}{\delta_2} [x_u + \delta_2 Q(v; \lambda)] \right]^{-\frac{1}{\lambda}} \right\} \\ &= \exp \left\{ - \left[1 + \frac{\lambda x_{1-u}}{\delta_1} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} - \exp \left\{ - \left[1 + \frac{\lambda x_u}{\delta_1} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} \\ &\quad - \exp \left\{ - \left[1 + \frac{\lambda x_{1-u}}{\delta_2} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} + \exp \left\{ - \left[1 + \frac{\lambda x_u}{\delta_2} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} \end{aligned}$$

Desde de que $X \sim N(0, 1)$, tem-se que $x_u < 0$ para $0 < u < 0.5$, e $x_{1-u} > 0$, em que $x_u = -x_{1-u}$, então

$$\begin{aligned} QS_{Y_1}(u) - QS_{Y_2}(u) &= \exp \left\{ - \left[1 + \frac{\lambda x_{1-u}}{\delta_1} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} - \exp \left\{ - \left[1 - \frac{\lambda x_{1-u}}{\delta_1} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} \\ &\quad - \exp \left\{ - \left[1 + \frac{\lambda x_{1-u}}{\delta_2} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} + \exp \left\{ - \left[1 - \frac{\lambda x_{1-u}}{\delta_2} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} \\ &= \exp \left\{ - \left[1 + \frac{\lambda x_{1-u}}{\delta_1} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} - \exp \left\{ - \left[1 + \frac{\lambda x_{1-u}}{\delta_2} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} \\ &\quad - \exp \left\{ - \left[1 - \frac{\lambda x_{1-u}}{\delta_1} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} + \exp \left\{ - \left[1 - \frac{\lambda x_{1-u}}{\delta_2} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} \leq 0, \end{aligned}$$

logo tem-se que

$$\begin{aligned} \exp \left\{ - \left[1 + \frac{\lambda x_{1-u}}{\delta_1} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} &\leq \exp \left\{ - \left[1 + \frac{\lambda x_{1-u}}{\delta_2} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\}, \text{ e} \\ \exp \left\{ - \left[1 - \frac{\lambda x_{1-u}}{\delta_2} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} &\leq \exp \left\{ - \left[1 - \frac{\lambda x_{1-u}}{\delta_1} + \lambda Q(v; \lambda) \right]^{-\frac{1}{\lambda}} \right\} \end{aligned}$$

se e só se $\delta_2 \leq \delta_1$. Portanto, $QS_{Y_1}(u) \leq QS_{Y_2}(u)$ para $0 < u < 0.5$, isto é $Y_1 \leq_{QS} Y_2$, se e só se $\delta_2 \leq \delta_1$. \square

Tendo demonstrado que o parâmetro δ é o parâmetro de dispersão do modelo de regressão Normal-VEG, e assumindo que ela varia entre as observações, então ele é relacionado a covariáveis assim como a mediana, e por isso são modelados conjuntamente.

3.2 Modelo de Regressão Normal-VEG

Seja $\mathbf{Y} = (y_1, y_2, \dots, y_n)$ um vetor que contém n realizações independentes da variável resposta Y , em que cada $Y_i \sim \text{Normal} - \text{VEG}(v_i, \delta_i, \lambda)$ com função densidade de probabilidade da forma

$$f_Y(y_i) = \frac{\delta_i \phi(\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)])}{y_i [-\log(y_i)]^{\lambda+1}}, \quad y_i \in (0, 1), \quad (3.4)$$

em que v_i e δ_i estão relacionadas com parâmetros desconhecidos e covariáveis mediante uma estrutura semelhante à empregada em modelos lineares generalizados, isto é

$$g_1(v_i) = \eta_{1i} = \mathbf{x}_i^\top \boldsymbol{\beta}$$

$$g_2(\delta_i) = \eta_{2i} = \mathbf{s}_i^\top \boldsymbol{\tau},$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ e $\boldsymbol{\tau} = (\tau_1, \dots, \tau_q)^\top$ são os vetores dos parâmetros desconhecidos da regressão, em que $\boldsymbol{\beta} \in \mathbb{R}^p$ e $\boldsymbol{\tau} \in \mathbb{R}^q$, com $p + q < n$, η_{1i} e η_{2i} são os preditores lineares e $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip})$ e $\mathbf{s}_i^\top = (s_{i1}, \dots, s_{iq})$ são as observações de p e q covariáveis conhecidas. Além disso, é assumido que as matrizes $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ e $\mathbf{S} = (\mathbf{s}_1, \dots, \mathbf{s}_n)^\top$ são de posto p

e q , respectivamente. As funções de ligação para a mediana e o parâmetro de dispersão são $g_1 : (0, 1) \rightarrow \mathbb{R}$ e $g_2 : (0, \infty) \rightarrow \mathbb{R}$, respectivamente, em que elas têm que ser estritamente monótonas e duplamente diferenciáveis. Na Tabela 9 são apresentadas algumas das funções de ligação para a mediana e para a dispersão.

Tabela 9 – Funções de ligação para a mediana e para o parâmetro de dispersão

Ligação para a mediana		Ligação para a dispersão	
Logito	$g_1(v) = \log\left(\frac{v}{1-v}\right)$	Logarítmica	$g_2(\delta) = \log(\delta)$
Probit	$g_1(v) = \Phi^{-1}(v)$	Raiz quadrada	$g_2(\delta) = \sqrt{\delta}$
Complemento log-log	$g_1(v) = \log[-\log(1-v)]$	Identidade	$g_2(\delta) = \delta$
Log-log	$g_1(v) = -\log[-\log(v)]$		
Cauchy	$g_1(v) = \tan[\pi(v-0.5)]$		

3.3 Inferência

Nesta seção é apresentado os procedimentos inferenciais para os parâmetros do modelo de regressão Normal-VEG, baseado no método de máxima verossimilhança.

Dado uma amostra observada de $(y_i, x_{i1}, \dots, x_{ip}, s_{i1}, \dots, s_{iq})$ para $i = 1, \dots, n$ e seja $\boldsymbol{\varphi} = (\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ o vetor de parâmetros desconhecidos, a função log-verossimilhança de $\boldsymbol{\varphi}$ é dado por

$$\ell(\boldsymbol{\varphi}) = \sum_{i=1}^n \log[f_Y(y_i)] = \sum_{i=1}^n \ell_i(\boldsymbol{\varphi}),$$

em que

$$\ell_i(\boldsymbol{\varphi}) = \log(\delta_i) + \log\{\phi(\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)])\} - \log(y_i) - \log[-\log(y_i)]^{\lambda+1},$$

em que $v_i = g_1^{-1}(\eta_{1i})$ é função de $\boldsymbol{\beta}$ e $\delta_i = g_2^{-1}(\eta_{2i})$ é função de $\boldsymbol{\tau}$. Os estimadores de máxima verossimilhança de $\boldsymbol{\varphi}$ são obtidas ao maximizar a função log-verossimilhança.

3.3.1 Função Escore

A função escore é obtida ao calcular a primeira derivada da função log-verossimilhança em relação a cada parâmetro do modelo (ver Anexo B). Neste caso, tem-se $p + q + 1$ parâmetros a estimar. O vetor escore é dado por:

$$\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \left(\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)^\top, \mathbf{U}_{\boldsymbol{\tau}}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)^\top, U_{\lambda}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) \right)^\top,$$

em que os componentes do vetor $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ têm a forma

$$\frac{\partial \ell(\boldsymbol{\varphi})}{\partial \beta_r} = \sum_{i=1}^n \frac{\delta_i w_i}{v_i [-\log(v_i)]^{\lambda+1}} \frac{x_{ir}}{\dot{g}_1(v_i)}, \quad r = 1, \dots, p.$$

Matricialmente é dada por

$$\mathbf{U}_\beta(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \mathbf{X}^\top \boldsymbol{\Delta} \mathbf{W} \mathbf{G}_1 \mathbf{V},$$

em que \mathbf{X} é a matriz modelo em relação a $\boldsymbol{\beta}$, $\boldsymbol{\Delta} = \text{diag}\{\delta_i\}$, $\mathbf{W} = \text{diag}\{w_i\}$, $\mathbf{G}_1 = \text{diag}\left\{\frac{1}{g_1(v_i)}\right\}$, $\mathbf{V} = (V_1, \dots, V_n)^\top$, sendo $V_i = \frac{1}{v_i[-\log(v_i)]^{\lambda+1}}$, para $i = 1, \dots, n$.

Da mesma forma os componentes de $\mathbf{U}_\tau(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ são dados por

$$\frac{\partial \ell(\boldsymbol{\varphi})}{\partial \tau_k} = \sum_{i=1}^n \left(\frac{1}{\delta_i} - \frac{w_i^2}{\delta_i} \right) \frac{s_{ik}}{g_2(\delta_i)}, \quad k = 1, \dots, q.$$

Matricialmente pode-se expressar como

$$\mathbf{U}_\tau(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \mathbf{S}^\top \mathbf{G}_2 \mathbf{T},$$

em que \mathbf{S} matriz modelo em relação a $\boldsymbol{\tau}$, $\mathbf{G}_2 = \text{diag}\left\{\frac{1}{g_2(\delta_i)}\right\}$, $\mathbf{T} = (T_1, \dots, T_n)^\top$, sendo $T_i = \frac{1}{\delta_i} - \frac{w_i^2}{\delta_i}$. Por último, tem-se que

$$\begin{aligned} U_\lambda(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) &= \sum_{i=1}^n \frac{w_i \delta_i}{\lambda} \left\{ [-\log(y_i)]^{-\lambda} \left\{ \frac{1}{\lambda} + \log[-\log(y_i)] \right\} - [-\log(v_i)]^{-\lambda} \left\{ \frac{1}{\lambda} + \log[-\log(v_i)] \right\} \right\} \\ &\quad - \sum_{i=1}^n \log[-\log(y_i)]. \end{aligned}$$

Os estimadores de máxima verossimilhança $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\tau}}$, $\hat{\lambda}$ são obtidos resolvendo os sistemas de equações

$$\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\tau}}, \hat{\lambda}) = \mathbf{0}_p, \quad \mathbf{U}_\tau(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\tau}}, \hat{\lambda}) = \mathbf{0}_q \quad \text{e} \quad U_\lambda(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\tau}}, \hat{\lambda}) = 0.$$

No entanto, as equações não têm soluções explícitas, mas podem ser resolvidas usando o algoritmo de Newton Raphson e, assim, podem-se obter os EMV dos parâmetros do modelo Normal-VEG.

Por outro lado, para n grande e sob certas condições de regularidade, tem-se que $\hat{\boldsymbol{\varphi}}$ segue uma distribuição assintótica normal, isto é

$$\hat{\boldsymbol{\varphi}} \sim N_{p+q+1}(\boldsymbol{\varphi}, \mathbf{I}(\boldsymbol{\varphi})^{-1}),$$

em que $\mathbf{I}(\boldsymbol{\varphi})$ é a matriz de informação de Fisher. O cálculo dessa matriz não é trivial, porém pelas propriedades assintóticas ela pode ser aproximada pela matriz de informação observada $\mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$, avaliada em $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\tau}}, \hat{\lambda}$. A matriz hessiana de ordem $(p+q+1)$ é dada por

$$\mathbf{J}_n(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \begin{bmatrix} \mathbf{J}_{\beta\beta} & \mathbf{J}_{\beta\tau} & \mathbf{J}_{\beta\lambda} \\ \mathbf{J}_{\beta\tau}^\top & \mathbf{J}_{\tau\tau} & \mathbf{J}_{\tau\lambda} \\ \mathbf{J}_{\beta\lambda}^\top & \mathbf{J}_{\tau\lambda}^\top & \mathbf{J}_{\lambda\lambda} \end{bmatrix},$$

cujos componentes são apresentados no Anexo C. Da normalidade assintótica dos EMV, é possível construir intervalos e regiões de confiança assintóticos de $100(1-\alpha)\%$ de confiança para os parâmetros do modelo de regressão Normal-VEG

$$(\hat{\boldsymbol{\beta}}_r - z_{1-\alpha/2} EP(\hat{\boldsymbol{\beta}}_r); \hat{\boldsymbol{\beta}}_r + z_{1-\alpha/2} EP(\hat{\boldsymbol{\beta}}_r)),$$

$$(\hat{\tau}_k - z_{1-\alpha/2} EP(\hat{\tau}_k); \hat{\tau}_k + z_{1-\alpha/2} EP(\hat{\tau}_k)) \text{ e} \\ (\hat{\lambda} - z_{1-\alpha/2} EP(\hat{\lambda}); \hat{\lambda} + z_{1-\alpha/2} EP(\hat{\lambda}))$$

para $r = 1, \dots, p$ e $k = 1, \dots, q$ respectivamente, em que $EP(\cdot)$ é o erro padrão assintótico, a qual é obtida pela raiz quadrada da diagonal da matriz $-\mathbf{J}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\tau}}, \hat{\lambda})^{-1}$ correspondente a cada parâmetro, $0 < \alpha < 1/2$ e $z_{1-\alpha/2}$ é o quantil $1 - \alpha/2$ da distribuição normal padrão.

3.3.2 Teste da razão de verossimilhança

Seja uma amostra aleatória (y_1, \dots, y_n) com função densidade de probabilidade em 3.4. O interesse é testar $H_0 : \boldsymbol{\varphi} = \mathbf{0}$, contra $H_1 : \boldsymbol{\varphi} \neq \mathbf{0}$, então a estatística do teste da razão de verossimilhança para testar estas hipóteses é dada por:

$$RV = 2[\ell(\hat{\boldsymbol{\varphi}}) - \ell(\tilde{\boldsymbol{\varphi}})],$$

em que $\ell(\hat{\boldsymbol{\varphi}})$ é a logaritmo da função verossimilhança aplicada no EMV de $\boldsymbol{\varphi}$ sob H_1 , e $\ell(\tilde{\boldsymbol{\varphi}})$ é o logaritmo da função verossimilhança aplicada no EMV de $\boldsymbol{\varphi}$ sob H_0 . Sob H_0 a estatística de razão de verossimilhanças (RV), tem distribuição assintótica qui-quadrado com $p + q + 1$ graus de liberdade (χ_{p+q+1}^2). Rejeitamos a hipótese H_0 para valores da estatística superior a $\chi_{p+q+1, \alpha}^2$, em que α é o nível de significância.

No caso em que tenha-se $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2)$, e queira-se testar $H_0 : \boldsymbol{\varphi}_2 = \mathbf{0}$, contra $H_1 : \boldsymbol{\varphi}_2 \neq \mathbf{0}$, então a estatística da razão de verossimilhança definida como acima

$$RV = 2[\ell(\hat{\boldsymbol{\varphi}}) - \ell(\tilde{\boldsymbol{\varphi}})],$$

sob H_0 tem distribuída assintótica χ_q^2 , em que q é a dimensão de $\boldsymbol{\varphi}_2$. Rejeitamos H_0 para valores da estatística superior a $\chi_{q, \alpha}^2$.

3.4 Estudo de simulação para o modelo de regressão Normal-VEG homocedástico

Um estudo de simulação é realizado com o objetivo de avaliar as propriedades assintóticas dos estimadores de máxima verossimilhança (EMV) dos parâmetros do modelo de regressão Normal-VEG homocedástico. Para o qual foram gerados amostras do par (y, x) , a covariável (x_i) foi gerada de uma $U(0, 1)$, e a variável resposta (y_i) foi gerada de $y_i \sim \text{Normal-VEG}(v_i, \delta, \lambda)$, onde

$$y_i = H\left(\frac{z_i}{\delta} + Q(v_i, \lambda)\right), \quad i = 1, \dots, n,$$

em que $H(\cdot)$ e $Q(\cdot)$ são dados em (1.6) e (2.1) respectivamente, z_i foi gerado de uma $N(0, 1)$, e a função de ligação que foi usada para relacionar a mediana da variável resposta (v_i) com a covariável e os parâmetros foi a logística isto é: $v_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$.

Assim, foram geradas amostras de tamanhos $n = 50, 100, 200, 400$ e 800 do modelo Normal-VEG homocedástico, com diferentes combinações de valores para os parâmetros como:

- $\beta_0 = 1, \beta_1 = -2, \delta = 2$ e $\lambda = 0,2$;
- $\beta_0 = -1, \beta_1 = 6, \delta = 2$ e $\lambda = -0,2$;
- $\beta_0 = -1, \beta_1 = 0,2, \delta = 10$ e $\lambda = -0,4$; e
- $\beta_0 = 1, \beta_1 = -0,2, \delta = 4$ e $\lambda = -0,6$.

Para cada combinação de tamanho de amostra e configuração dos parâmetros, foram simulados 1.000 amostras, e com as estimativas de máxima verossimilhança foram calculadas a média, a média do erro padrão (MEP), o desvio padrão (DP), o viés, a raiz quadrada do erro quadrático médio (REQM) e a probabilidade de cobertura (PC) que foram obtidas como no Capítulo 2. Os resultados são apresentados e discutidos como segue:

- Na Tabela 10 são apresentados os resultados da simulação do modelo Normal-VEG homocedástico, considerando $\beta_0 = 1, \beta_1 = -2, \delta = 2$ e $\lambda = 0,2$. Pode-se observar que as EMVs se aproximam dos verdadeiros valores dos parâmetros, e assim tornando um viés muito pequeno a medida que o tamanho da amostra aumenta. O erro padrão o desvio padrão e a raiz do erro quadrático médio das estimativas se aproximam de zero quando o tamanho da amostra aumenta. E a probabilidade de cobertura captura melhor a proporção das amostras que contém o verdadeiro valor do parâmetro isto é ao redor de 95%.
- Na Tabela 11 são apresentados os resultados da simulação considerando $\beta_0 = -1, \beta_1 = 6, \delta = 2$ e $\lambda = -0,2$. Para esta combinação de parâmetros os resultados são similares ao item anterior notando que apresenta melhores resultados para amostras de tamanho a partir de 200.
- Na Tabela 12 são apresentados os resultados da simulação considerando $\beta_0 = -1, \beta_1 = 0,2, \delta = 10$, e $\lambda = -0,4$. Para esta combinação de parâmetros os resultados são os esperados, obtendo boas estimativas assintoticamente com viés pequeno. Mas pode-se notar que para o valor do parâmetro $\delta = 10$ o viés é maior que o viés das outras estimativas, mas ainda assim pode-se observar que ele tem uma probabilidade de cobertura de 95% que é o esperado, com tudo não deixando de ter uma boa estimativa para esse parâmetro.
- Na Tabela 13 são apresentados os resultados da simulação considerando $\beta_0 = 1, \beta_1 = -0,2, \delta = 4$ e $\lambda = -0,6$. Para esta combinação de parâmetros, observa-se que para $\delta = 4$, as estimativas mostram um viés e a REQM um pouco alto, mas com uma probabilidade de cobertura ao redor de 95% para amostras maiores de 800.

Tabela 10 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG homocedástico com parâmetros $\beta_0 = 1$, $\beta_1 = -2$, $\delta = 2$ e $\lambda = 0,2$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	0,999	0,152	0,161	-0,001	0,161	0,927
	β_1	-1,998	0,322	0,342	0,002	0,341	0,934
	δ	2,095	0,260	0,279	0,095	0,295	0,953
	λ	0,202	0,188	0,188	0,002	0,188	0,948
100	β_0	1,001	0,108	0,110	0,001	0,110	0,934
	β_1	-2,004	0,228	0,230	-0,004	0,230	0,938
	δ	2,040	0,174	0,171	0,040	0,176	0,962
	λ	0,203	0,127	0,125	0,003	0,125	0,964
200	β_0	0,997	0,076	0,077	-0,003	0,077	0,947
	β_1	-1,994	0,162	0,163	0,006	0,163	0,949
	δ	2,019	0,121	0,120	0,019	0,122	0,955
	λ	0,201	0,088	0,086	0,001	0,086	0,952
400	β_0	0,998	0,054	0,055	-0,002	0,055	0,944
	β_1	-1,999	0,114	0,115	0,001	0,115	0,942
	δ	2,015	0,085	0,087	0,015	0,089	0,939
	λ	0,197	0,061	0,063	-0,003	0,063	0,944
800	β_0	0,999	0,038	0,038	-0,001	0,038	0,955
	β_1	-2,000	0,081	0,082	0,000	0,082	0,951
	δ	2,006	0,060	0,059	0,006	0,060	0,946
	λ	0,200	0,043	0,045	0,000	0,044	0,939

Tabela 11 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG homocedástico com parâmetros $\beta_0 = -1$, $\beta_1 = 6$, $\delta = 2$ e $\lambda = -0,2$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	-0,989	0,235	0,255	0,011	0,255	0,918
	β_1	5,984	0,495	0,526	-0,016	0,526	0,933
	δ	2,090	0,353	0,364	0,090	0,375	0,964
	λ	-0,221	0,061	0,052	-0,021	0,057	0,968
100	β_0	-0,994	0,165	0,169	0,006	0,169	0,940
	β_1	5,986	0,345	0,344	-0,014	0,344	0,945
	δ	2,057	0,238	0,249	0,057	0,255	0,948
	λ	-0,218	0,041	0,036	-0,018	0,041	0,959
200	β_0	-1,002	0,117	0,118	-0,002	0,118	0,948
	β_1	6,001	0,242	0,241	0,001	0,241	0,956
	δ	2,017	0,163	0,163	0,017	0,164	0,948
	λ	-0,211	0,028	0,025	-0,011	0,027	0,959
400	β_0	-0,996	0,082	0,083	0,004	0,083	0,945
	β_1	5,993	0,170	0,167	-0,007	0,167	0,946
	δ	2,006	0,114	0,116	0,006	0,116	0,950
	λ	-0,208	0,020	0,018	-0,008	0,019	0,954
800	β_0	-1,000	0,058	0,058	0,000	0,058	0,948
	β_1	6,001	0,119	0,122	0,001	0,122	0,949
	δ	2,002	0,080	0,077	0,002	0,077	0,960
	λ	-0,206	0,014	0,012	-0,006	0,013	0,956

Tabela 12 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG homocedástico com parâmetros $\beta_0 = -1$, $\beta_1 = 0,2$, $\delta = 10$ e $\lambda = -0,4$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	-0,998	0,045	0,047	0,002	0,047	0,931
	β_1	0,197	0,077	0,081	-0,003	0,081	0,931
	δ	10,889	3,228	3,438	0,889	3,549	0,969
	λ	-0,388	1,291	1,274	0,012	1,274	0,960
100	β_0	-0,998	0,032	0,033	0,002	0,033	0,947
	β_1	0,197	0,055	0,055	-0,003	0,056	0,932
	δ	10,360	2,073	2,152	0,360	2,181	0,952
	λ	-0,411	0,869	0,857	-0,011	0,856	0,961
200	β_0	-1,000	0,023	0,023	0,000	0,023	0,938
	β_1	0,200	0,039	0,039	0,000	0,039	0,949
	δ	10,190	1,401	1,419	0,190	1,431	0,951
	λ	-0,392	0,596	0,580	0,008	0,579	0,962
400	β_0	-1,000	0,016	0,017	0,000	0,017	0,942
	β_1	0,200	0,028	0,028	0,000	0,028	0,947
	δ	10,139	0,972	0,969	0,139	0,978	0,948
	λ	-0,381	0,416	0,407	0,019	0,407	0,944
800	β_0	-1,000	0,012	0,011	0,000	0,011	0,962
	β_1	0,200	0,020	0,019	0,000	0,019	0,951
	δ	10,061	0,678	0,678	0,061	0,681	0,950
	λ	-0,393	0,292	0,288	0,007	0,288	0,950

Tabela 13 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG homocedástico com parâmetros $\beta_0 = 1$, $\beta_1 = -0,2$, $\delta = 4$ e $\lambda = -0,6$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	1,015	0,163	0,168	0,015	0,169	0,934
	β_1	-0,207	0,273	0,287	-0,007	0,287	0,933
	δ	4,071	1,215	1,221	0,071	1,222	0,940
	λ	-0,544	0,243	0,235	0,056	0,241	0,961
100	β_0	1,010	0,116	0,118	0,010	0,118	0,953
	β_1	-0,210	0,194	0,197	-0,010	0,197	0,935
	δ	3,984	0,801	0,750	-0,016	0,750	0,956
	λ	-0,564	0,163	0,154	0,036	0,158	0,963
200	β_0	1,001	0,083	0,084	0,001	0,084	0,936
	β_1	-0,195	0,138	0,139	0,005	0,139	0,948
	δ	3,956	0,543	0,504	-0,044	0,506	0,957
	λ	-0,576	0,111	0,102	0,024	0,105	0,967
400	β_0	1,002	0,058	0,060	0,002	0,060	0,937
	β_1	-0,199	0,097	0,097	0,001	0,097	0,950
	δ	3,955	0,378	0,361	-0,045	0,364	0,944
	λ	-0,581	0,077	0,073	0,019	0,076	0,953
800	β_0	1,000	0,041	0,042	0,000	0,042	0,953
	β_1	-0,199	0,069	0,070	0,001	0,070	0,950
	δ	3,947	0,264	0,247	-0,053	0,253	0,950
	λ	-0,586	0,054	0,052	0,014	0,053	0,949

3.5 Estudo de simulação para o modelo de regressão Normal-VEG heterocedásticos

Nesta Seção um estudo de simulação foi realizado com o objetivo de avaliar as propriedades assintóticas dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão Normal-VEG heterocedástico apresentada no Capítulo 3. Assim como no caso homocedástico foi gerado amostras do par (y, x) , mas desta vez amostras da variável resposta (y_i) foram geradas levando em conta dispersão não constante, isto é, $y_i \sim \text{Normal-VEG}(v_i, \delta_i, \lambda)$, onde

$$y_i = H\left(\frac{z_i}{\delta_i} + Q(v_i, \lambda)\right), \quad i = 1, \dots, n,$$

em que $H(\cdot)$ e $Q(\cdot)$ são dados em (1.6) e (2.1) respectivamente, z_i foi gerado de uma $N(0, 1)$. A variável explicativa (x_i) foi gerada de uma $U(0, 1)$. As funções de ligação utilizadas foram logística para a mediana e logarítmica para a dispersão, isto é: $v_i = e^{\beta_0 + \beta_1 x_i} / (1 + e^{\beta_0 + \beta_1 x_i})$ e $\delta_i = \exp\{\tau_0 + \tau_1 x_i\}$.

Neste estudo foram geradas amostras de tamanho $n = 50, 100, 200, 400$ e 800 do modelo de regressão Normal-VEG heterocedástico com diferentes configurações de parâmetros como:

- $\beta_0 = 1, \beta_1 = -2, \tau_0 = 2, \tau_1 = 1$ e $\lambda = -0,2$;
- $\beta_0 = 1, \beta_1 = -2, \tau_0 = 2, \tau_1 = 1$ e $\lambda = -0,7$;
- $\beta_0 = -4, \beta_1 = -2, \tau_0 = 2, \tau_1 = 1$ e $\lambda = 0,7$;
- $\beta_0 = -3, \beta_1 = -2, \tau_0 = 1, \tau_1 = 1$ e $\lambda = 0,4$;
- $\beta_0 = -6, \beta_1 = 8, \tau_0 = 0,2, \tau_1 = 1$ e $\lambda = -0,2$;
- $\beta_0 = -1, \beta_1 = 2, \tau_0 = 1, \tau_1 = 0,4$ e $\lambda = 0,5$; e
- $\beta_0 = -5, \beta_1 = -1, \tau_0 = 1, \tau_1 = 2$ e $\lambda = 0,1$.

Para cada combinação de tamanho de amostra e configuração dos parâmetros foram simulados 1.000 amostras, e por meio das estimativas de máxima verossimilhança dos parâmetros do modelo foram calculadas a média, a média do erro padrão (MEP), o desvio padrão (DP), o viés, a raiz quadrada do erro quadrático médio (REQM) e a probabilidade de cobertura (PC). Os resultados são discutidos como segue:

- Na Tabela 14 são apresentados os resultados da simulação para $\beta_0 = 1, \beta_1 = -2, \tau_0 = 2, \tau_1 = 1$ e $\lambda = -0,2$. Pode-se ver que as estimativas dos parâmetros se aproximam dos verdadeiros valores dos parâmetros em média a medida que o tamanho da amostra aumenta. O viés vai tornando-se mais pequeno, a média do erro padrão, o desvio padrão e a raiz do erro quadrático médio pequenos, a medida que o tamanho da amostra aumenta. Mas

notando que para os parâmetros τ_0 , τ_1 e $\lambda = -0,2$ não tão rapidamente como para os outros parâmetros com essa combinação de parâmetros.

- Na Tabela 15 são apresentados os resultados da simulação para $\beta_0 = 1$, $\beta_1 = -2$, $\tau_0 = 2$, $\tau_1 = 1$ e $\lambda = -0,7$ na qual pode-se observar que as estimativas para $\beta_0 = 1$ e $\beta_1 = -2$, o viés é pequeno para pequenas amostras tornando-se zero para grandes amostras, o que não acontece tão rapidamente para os parâmetros $\tau_0 = 2$, $\tau_1 = 1$ e $\lambda = -0,7$, para essa combinação de parâmetros.
- Na Tabela 16 são apresentados os resultados da simulação para $\beta_0 = -4$, $\beta_1 = -2$, $\tau_0 = 2$, $\tau_1 = 1$ e $\lambda = 0,7$. Notamos que o viés diminui para todos os parâmetros à medida que o tamanho da amostra aumenta. Podemos observar que para um tamanho de amostra 400 as estimativas possuem boas propriedades para este conjunto de valores dos parâmetros.
- Na Tabela 17 são apresentados os resultados da simulação para $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$ e $\lambda = 0,4$, para o qual temos resultados parecidos ao item anterior, mostrando com isso a convergência assintótica das EMVs dos parâmetros para o verdadeiro valor.
- Na Tabela 18 são apresentados os resultados da simulação para $\beta_0 = -6$, $\beta_1 = 8$, $\tau_0 = 0,2$, $\tau_1 = 1$ e $\lambda = -0,2$, em que para este conjunto de parâmetros as estimativas vão-se tornando melhores para tamanhos de amostra maiores que 200, com medidas de precisão muito pequenas e com probabilidade de cobertura ao redor de 95%.
- Na Tabela 19 são apresentados os resultados da simulação para $\beta_0 = -1$, $\beta_1 = 2$, $\tau_0 = 1$, $\tau_1 = 0,4$ e $\lambda = 0,5$. Aqui vemos de novo o bom desempenho das estimativas de máxima verossimilhança que mostram viés pequeno perto de zero e erro padrão, desvio padrão e raiz do erro quadrático médio se aproximando de zero, o qual nota-se melhor a partir de um tamanho de amostra 400.
- Na Tabela 20 são apresentados os resultados da simulação para $\beta_0 = -5$, $\beta_1 = -1$, $\tau_0 = 1$, $\tau_1 = 2$ e $\lambda = 0,1$. Aqui também para essa combinação de parâmetros as EMVs dos parâmetros se aproximam do verdadeiro valor dos parâmetros, com um viés cada vez mais pequeno à medida que o tamanho da amostra aumenta. Notando que para $\lambda = 0,1$, o viés dele se torna zero para o tamanho de amostra $n = 200$, o mesmo acontece com o parâmetro $\tau_1 = 2$ quando $n = 800$, e tornando-se muito perto de zero para os demais parâmetros.

Com tudo, provando com esse estudo de simulação que as EMVs dos parâmetros do modelo de regressão homocedástico e heterocedástico possuem boas propriedades assintóticas.

Tabela 14 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = 1$, $\beta_1 = -2$, $\tau_0 = 2$, $\tau_1 = 1$ e $\lambda = -0,2$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	1,003	0,044	0,046	0,003	0,046	0,927
	β_1	-2,004	0,063	0,067	-0,004	0,067	0,926
	τ_0	1,963	1,468	1,410	-0,037	1,410	0,973
	τ_1	1,107	1,915	1,835	0,107	1,838	0,972
	λ	-0,119	1,304	1,246	0,081	1,248	0,972
100	β_0	1,002	0,031	0,032	0,002	0,032	0,951
	β_1	-2,003	0,045	0,045	-0,003	0,045	0,951
	τ_0	1,998	0,962	0,928	-0,002	0,928	0,959
	τ_1	1,026	1,255	1,220	0,026	1,219	0,962
	λ	-0,170	0,854	0,826	0,030	0,826	0,963
200	β_0	1,000	0,022	0,023	0,000	0,023	0,939
	β_1	-2,000	0,032	0,032	0,000	0,032	0,939
	τ_0	1,984	0,650	0,629	-0,016	0,629	0,962
	τ_1	1,032	0,848	0,822	0,032	0,823	0,962
	λ	-0,173	0,577	0,561	0,027	0,561	0,961
400	β_0	1,000	0,016	0,016	0,000	0,016	0,950
	β_1	-2,000	0,023	0,022	0,000	0,022	0,948
	τ_0	1,976	0,449	0,441	-0,024	0,441	0,950
	τ_1	1,042	0,586	0,574	0,042	0,575	0,956
	λ	-0,176	0,399	0,395	0,024	0,396	0,949
800	β_0	1,000	0,011	0,011	0,000	0,011	0,962
	β_1	-2,000	0,016	0,016	0,000	0,016	0,952
	τ_0	1,990	0,313	0,306	-0,010	0,306	0,957
	τ_1	1,017	0,409	0,400	0,017	0,400	0,957
	λ	-0,188	0,278	0,273	0,012	0,273	0,958

Tabela 15 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = 1$, $\beta_1 = -2$, $\tau_0 = 2$, $\tau_1 = 1$ e $\lambda = -0,7$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	1,006	0,060	0,063	0,006	0,063	0,930
	β_1	-2,007	0,079	0,083	-0,007	0,083	0,935
	τ_0	1,911	1,065	1,014	-0,089	1,018	0,976
	τ_1	1,177	1,400	1,331	0,177	1,342	0,978
	λ	-0,576	0,925	0,878	0,124	0,886	0,975
100	β_0	1,004	0,043	0,043	0,004	0,043	0,960
	β_1	-2,005	0,056	0,056	-0,005	0,057	0,958
	τ_0	1,958	0,688	0,655	-0,042	0,656	0,965
	τ_1	1,078	0,905	0,870	0,078	0,873	0,961
	λ	-0,635	0,596	0,569	0,065	0,572	0,965
200	β_0	1,001	0,031	0,031	0,001	0,031	0,946
	β_1	-2,001	0,040	0,040	-0,001	0,040	0,945
	τ_0	1,966	0,460	0,443	-0,034	0,444	0,962
	τ_1	1,055	0,607	0,585	0,055	0,587	0,957
	λ	-0,657	0,398	0,387	0,043	0,389	0,966
400	β_0	1,000	0,022	0,022	0,000	0,022	0,950
	β_1	-2,000	0,028	0,028	0,000	0,028	0,960
	τ_0	1,968	0,316	0,304	-0,032	0,306	0,957
	τ_1	1,053	0,416	0,401	0,053	0,404	0,957
	λ	-0,669	0,272	0,267	0,031	0,269	0,958
800	β_0	1,000	0,015	0,015	0,000	0,015	0,967
	β_1	-2,000	0,020	0,019	0,000	0,019	0,955
	τ_0	1,985	0,219	0,211	-0,015	0,212	0,961
	τ_1	1,023	0,289	0,280	0,023	0,281	0,959
	λ	-0,684	0,188	0,183	0,016	0,183	0,965

Tabela 16 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = -4$, $\beta_1 = -2$, $\tau_0 = 2$, $\tau_1 = 1$ e $\lambda = 0,7$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	-4,017	0,395	0,411	-0,017	0,411	0,923
	β_1	-1,998	0,620	0,654	0,002	0,653	0,930
	τ_0	1,966	0,740	0,701	-0,034	0,701	0,961
	τ_1	0,961	0,423	0,422	-0,039	0,424	0,958
	λ	0,637	0,492	0,469	-0,063	0,473	0,969
100	β_0	-4,005	0,280	0,286	-0,005	0,286	0,949
	β_1	-2,012	0,437	0,444	-0,012	0,444	0,941
	τ_0	1,981	0,488	0,479	-0,019	0,479	0,965
	τ_1	0,968	0,282	0,287	-0,032	0,288	0,953
	λ	0,664	0,323	0,314	-0,036	0,315	0,963
200	β_0	-4,010	0,200	0,203	-0,010	0,203	0,938
	β_1	-1,990	0,311	0,316	0,010	0,316	0,943
	τ_0	1,991	0,332	0,323	-0,009	0,323	0,957
	τ_1	0,987	0,194	0,191	-0,013	0,191	0,945
	λ	0,684	0,218	0,210	-0,016	0,211	0,957
400	β_0	-4,007	0,142	0,144	-0,007	0,144	0,942
	β_1	-1,994	0,219	0,219	0,006	0,219	0,949
	τ_0	1,992	0,230	0,229	-0,008	0,229	0,947
	τ_1	1,004	0,135	0,135	0,004	0,135	0,949
	λ	0,693	0,151	0,148	-0,007	0,148	0,953
800	β_0	-4,005	0,100	0,098	-0,005	0,098	0,957
	β_1	-1,996	0,155	0,153	0,004	0,153	0,954
	τ_0	1,996	0,160	0,161	-0,004	0,161	0,948
	τ_1	0,998	0,095	0,095	-0,002	0,095	0,947
	λ	0,695	0,105	0,105	-0,005	0,105	0,953

Tabela 17 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$ e $\lambda = 0,4$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	-3,023	0,495	0,514	-0,023	0,514	0,928
	β_1	-1,997	0,774	0,813	0,003	0,813	0,931
	τ_0	1,014	0,435	0,416	0,014	0,416	0,959
	τ_1	0,968	0,408	0,407	-0,032	0,408	0,958
	λ	0,364	0,320	0,306	-0,036	0,308	0,968
100	β_0	-3,007	0,351	0,359	-0,007	0,359	0,946
	β_1	-2,014	0,545	0,554	-0,014	0,554	0,941
	τ_0	1,008	0,289	0,288	0,008	0,288	0,959
	τ_1	0,971	0,272	0,277	-0,029	0,279	0,949
	λ	0,380	0,210	0,204	-0,020	0,205	0,962
200	β_0	-3,013	0,250	0,254	-0,013	0,255	0,941
	β_1	-1,988	0,388	0,393	0,012	0,393	0,944
	τ_0	1,003	0,198	0,195	0,003	0,195	0,949
	τ_1	0,989	0,188	0,185	-0,011	0,185	0,950
	λ	0,391	0,142	0,137	-0,009	0,138	0,957
400	β_0	-3,008	0,178	0,180	-0,008	0,180	0,947
	β_1	-1,994	0,274	0,273	0,006	0,273	0,949
	τ_0	0,998	0,138	0,139	-0,002	0,138	0,951
	τ_1	1,004	0,131	0,131	0,004	0,131	0,946
	λ	0,396	0,098	0,097	-0,004	0,097	0,954
800	β_0	-3,007	0,126	0,123	-0,007	0,123	0,957
	β_1	-1,993	0,194	0,191	0,007	0,192	0,953
	τ_0	1,000	0,096	0,097	0,000	0,097	0,944
	τ_1	0,998	0,092	0,092	-0,002	0,092	0,951
	λ	0,398	0,069	0,069	-0,002	0,069	0,953

Tabela 18 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = -6$, $\beta_1 = 8$, $\tau_0 = 0,2$, $\tau_1 = 1$ e $\lambda = -0,2$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	-5,967	0,457	0,480	0,033	0,481	0,928
	β_1	7,963	0,581	0,619	-0,037	0,620	0,927
	τ_0	0,315	0,579	0,574	0,115	0,585	0,958
	τ_1	0,900	0,950	0,937	-0,100	0,942	0,959
	λ	-0,163	0,239	0,237	0,037	0,239	0,951
100	β_0	-5,979	0,326	0,327	0,021	0,327	0,951
	β_1	7,972	0,413	0,415	-0,028	0,416	0,951
	τ_0	0,251	0,383	0,380	0,051	0,384	0,950
	τ_1	0,954	0,629	0,614	-0,046	0,616	0,957
	λ	-0,187	0,157	0,154	0,013	0,155	0,959
200	β_0	-5,995	0,232	0,231	0,005	0,231	0,946
	β_1	7,995	0,294	0,294	-0,005	0,294	0,947
	τ_0	0,229	0,262	0,256	0,029	0,257	0,957
	τ_1	0,970	0,430	0,421	-0,030	0,422	0,957
	λ	-0,191	0,106	0,103	0,009	0,103	0,955
400	β_0	-6,001	0,164	0,166	-0,001	0,165	0,949
	β_1	8,002	0,208	0,207	0,002	0,207	0,953
	τ_0	0,213	0,182	0,188	0,013	0,188	0,955
	τ_1	0,991	0,299	0,310	-0,009	0,310	0,956
	λ	-0,195	0,074	0,075	0,005	0,075	0,953
800	β_0	-6,003	0,116	0,112	-0,003	0,112	0,958
	β_1	8,005	0,147	0,143	0,005	0,143	0,955
	τ_0	0,208	0,127	0,128	0,008	0,129	0,951
	τ_1	0,991	0,209	0,211	-0,009	0,211	0,945
	λ	-0,197	0,052	0,051	0,003	0,051	0,955

Tabela 19 – Resultados baseados em 1000 simulações do modelo de regressão heterocedástico com parâmetros $\beta_0 = -1$, $\beta_1 = 2$, $\tau_0 = 1$, $\tau_1 = 0,4$ e $\lambda = 0,5$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	-1,000	0,137	0,142	0,000	0,142	0,940
	β_1	1,998	0,178	0,186	-0,002	0,186	0,934
	τ_0	1,033	0,276	0,270	0,033	0,271	0,953
	τ_1	0,461	0,803	0,766	0,061	0,768	0,959
	λ	0,461	0,484	0,460	-0,039	0,461	0,970
100	β_0	-0,999	0,098	0,099	0,001	0,098	0,957
	β_1	1,996	0,126	0,127	-0,004	0,127	0,959
	τ_0	1,020	0,185	0,188	0,020	0,189	0,952
	τ_1	0,423	0,530	0,516	0,023	0,516	0,961
	λ	0,476	0,317	0,306	-0,024	0,307	0,965
200	β_0	-1,001	0,070	0,070	-0,001	0,070	0,943
	β_1	2,001	0,090	0,091	0,001	0,091	0,940
	τ_0	1,009	0,128	0,127	0,009	0,127	0,954
	τ_1	0,410	0,361	0,354	0,010	0,354	0,953
	λ	0,490	0,214	0,206	-0,010	0,206	0,956
400	β_0	-1,002	0,049	0,050	-0,002	0,050	0,946
	β_1	2,002	0,064	0,063	0,002	0,063	0,956
	τ_0	1,002	0,089	0,091	0,002	0,091	0,955
	τ_1	0,412	0,251	0,254	0,012	0,254	0,950
	λ	0,498	0,147	0,145	-0,002	0,145	0,955
800	β_0	-1,001	0,035	0,034	-0,001	0,034	0,966
	β_1	2,002	0,045	0,044	0,002	0,044	0,958
	τ_0	1,002	0,063	0,063	0,002	0,063	0,943
	τ_1	0,403	0,175	0,176	0,003	0,176	0,945
	λ	0,498	0,103	0,102	-0,002	0,102	0,954

Tabela 20 – Resultados baseados em 1000 simulações do modelo de regressão Normal-VEG heterocedástico com parâmetros $\beta_0 = -5$, $\beta_1 = -1$, $\tau_0 = 1$, $\tau_1 = 2$ e $\lambda = 0,1$.

n	Parâmetro	Média	MEP	DP	Viés	REQM	PC
50	β_0	-4,992	0,360	0,374	0,008	0,374	0,933
	β_1	-1,012	0,457	0,478	-0,012	0,478	0,937
	τ_0	1,028	1,073	1,014	0,028	1,014	0,973
	τ_1	1,997	0,399	0,396	-0,003	0,396	0,961
	λ	0,087	0,651	0,617	-0,013	0,617	0,974
100	β_0	-4,992	0,257	0,258	0,008	0,258	0,958
	β_1	-1,014	0,325	0,326	-0,014	0,326	0,960
	τ_0	1,017	0,696	0,670	0,017	0,670	0,964
	τ_1	1,983	0,266	0,271	-0,017	0,271	0,958
	λ	0,091	0,421	0,403	-0,009	0,403	0,966
200	β_0	-5,001	0,183	0,184	-0,001	0,184	0,944
	β_1	-0,999	0,231	0,233	0,001	0,233	0,944
	τ_0	1,013	0,468	0,457	0,013	0,457	0,958
	τ_1	1,994	0,184	0,181	-0,006	0,181	0,950
	λ	0,100	0,282	0,273	0,000	0,273	0,955
400	β_0	-5,002	0,130	0,131	-0,002	0,131	0,949
	β_1	-0,997	0,163	0,162	0,003	0,162	0,951
	τ_0	1,011	0,322	0,320	0,011	0,320	0,956
	τ_1	2,009	0,128	0,129	0,009	0,129	0,942
	λ	0,105	0,194	0,190	0,005	0,190	0,954
800	β_0	-5,003	0,092	0,089	-0,003	0,089	0,966
	β_1	-0,996	0,116	0,111	0,004	0,111	0,959
	τ_0	1,006	0,224	0,222	0,006	0,222	0,958
	τ_1	2,000	0,090	0,091	0,000	0,091	0,945
	λ	0,102	0,135	0,134	0,002	0,133	0,962

3.6 Aplicação

Nesta Seção, para mostrar a utilidade do modelo de regressão Normal-VEG proposto, foi ajustado o modelo de regressão a um conjunto de dados de câncer colorretal introduzidos por Siegel, DeSantis e Jemal (2014). O câncer colorretal é o terceiro câncer mais comum e a terceira principal causa de morte por câncer em homens e mulheres nos Estados Unidos. As estimativas do Instituto Nacional do Câncer (NCI) para 2017 são de 135 mil novos casos nos EUA, 8% do total de novos casos de câncer para aquele ano. O NCI também informou que a estimativa de mortes para 2017 é de cerca de 50 mil, 8,4% de todas as mortes por câncer. A melhoria na tecnologia para detectar o câncer, o avanço nas técnicas de tratamento e a diminuição da exposição a fatores de risco (como tabagismo e consumo de carne), todos contribuíram para a redução da mortalidade por câncer colorretal de sete décadas atrás, quando foi a primeira causa de morte por câncer nos EUA.

O conjunto de dados provêm de 50 estados dos EUA entre os anos 2006 e 2010, da qual considerou-se a taxa de sobrevivência como a variável resposta. Após excluir estados com dados incompletos, foi obtido um subconjunto de $n = 220$ observações. O objetivo é avaliar se existem diferenças significativas na taxa de sobrevivência entre homens e mulheres, assim como também entre pacientes de diferentes raças. Para este propósito, as seguintes variáveis foram consideradas:

- y_i : taxa de sobrevivência, para $i = 1, \dots, 220$;
- x_{1i} : sexo (homem=0, mulher=1);
- x_{2i} : raça (não hispânico branco, não hispânico negro, hispânico).

Onde a taxa de sobrevivência foi definido como o complemento da taxa de mortalidade, em que a taxa de mortalidade é calculada pelos autores como:

$$\text{Taxa de mortalidade} = \frac{\text{número de casos de mortalidade de câncer colorretal}}{100.000 \text{ habitantes}},$$

então, a taxa de sobrevivência = $1 - \text{taxa de mortalidade}$. Todos os cálculos foram obtidos usando o software R Core Team (2015). Na Tabela 21 são apresentadas as medidas descritivas da taxa de sobrevivência de câncer colorretal.

Tabela 21 – Medidas descritivas da taxa de sobrevivência das pessoas com câncer colorretal nos EUA

Min.	1 Qu.	Mediana	Média	3 Qu.	Max.
0.6280	0.7880	0.8230	0.8185	0.8670	0.9530

A Figura 9 mostra o boxplot da taxa de sobrevivência por sexo e por raça, onde pode-se notar que a taxa de sobrevivência do câncer é diferente para homens e para mulheres, da mesma forma ocorre com as raças. Por exemplo note que a taxa de sobrevivência de pacientes de origem

hispanico é maior que os outros dois de origem não-hispanico. Para avaliar essa diferença, foi ajustado o modelo de regressão descrito na Seção 3.2 aos dados.

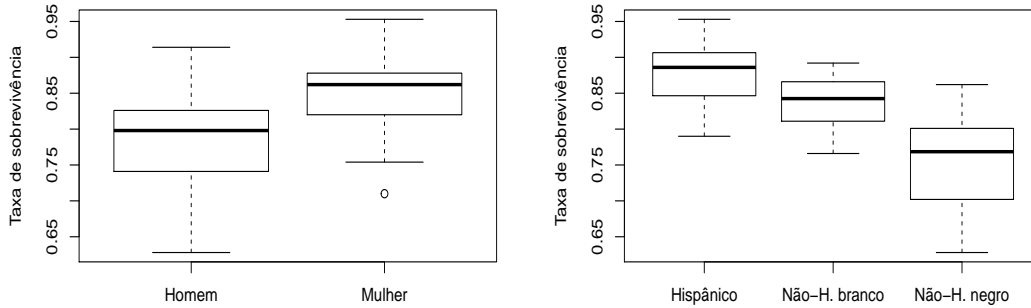


Figura 9 – Boxplot da taxa de sobrevivência de pacientes com câncer colorretal por sexo (lado esquerdo) e por raça (lado direito)

Assim sendo, é modelada a mediana da taxa de sobrevivência do paciente com câncer colorretal (v) e o parâmetro de dispersão (δ) com todas as covariáveis. Consideramos as seguintes funções de ligação: logística, probit, cauchy, complemento log-log, e log-log para a mediana e logarítmica para o parâmetro de dispersão, isto é:

$$\begin{aligned} g_1(v_i) &= \beta_0 + \beta_1 x_{1i} + \beta_{2_1} x_{2_1i} + \beta_{2_2} x_{2_2i}, \quad e \\ g_2(\delta_i) &= \log(\delta_i) = \tau_0 + \tau_1 x_{1i} + \tau_{2_1} x_{2_1i} + \tau_{2_2} x_{2_2i}, \quad i = 1, \dots, 220. \end{aligned} \quad (3.5)$$

Em que, dado que tem-se duas variáveis categóricas de dois e três níveis respectivamente, tem-se que:

$$x_{1i} = \begin{cases} 1, & \text{se mulher;} \\ 0, & \text{se homem} \end{cases},$$

$$x_{2_1i} = \begin{cases} 1, & \text{se não-hispanico branco;} \\ 0, & \text{outro caso} \end{cases} \quad e \quad x_{2_2i} = \begin{cases} 1, & \text{se não-hispanico negro;} \\ 0, & \text{outro caso.} \end{cases}$$

Os dados foram ajustados com o modelo de regressão Normal-VEG com todas as ligações mencionadas, e foi feita uma seleção de modelos usando os critérios AIC (BOZDOGAN, 1987) e BIC (WATANABE, 2013). Os valores dos critérios são apresentados na Tabela 22, da qual, de acordo com os critérios, o modelo que melhor ajusta esses dados é o modelo Normal-VEG com a ligação log-log o qual é denominado de modelo Normal-VEG-LL.

Com a finalidade de comparar modelos, os dados também foram ajustados com o modelo de regressão S_B de Johnson (com função de ligação log-log). Na Tabela 23 são apresentados as estimativas de máxima verossimilhança e seu respectivo erro padrão dos parâmetros dos

Tabela 22 – Critérios de comparação de modelos AIC e BIC do modelo Normal-VEG com diferentes funções de ligações para os dados de câncer colorretal

Função de ligação	Critério	
	AIC	BIC
Logito	-995,894	-965,351
Probit	-993,991	-963,448
Cauchy	-993,779	-963,237
Com-log-log	-990,552	-960,010
Log-log	-997,219	-966,677

modelos ajustados. Da qual pode-se ver que todos os parâmetros saíram significativos para ambos modelos. O qual significa que o efeito do sexo e da raça dos pacientes com câncer colorretal são significativos na taxa de sobrevivência mediana. Isto quer dizer que a mediana da taxa de sobrevivência difere entre homens e mulheres, assim como também difere entre pacientes de origem não-hispânico branco, não-hispânico negro e hispânico. Similarmente o efeito das covariáveis sexo e raça do paciente com câncer colorretal são significativos na dispersão do modelo. Isto quer dizer que existem diferenças significativas na dispersão entre os pacientes homens e mulheres, e entre os pacientes de origem hispânico e os outros dois de origem não-hispânico.

Tabela 23 – Estimativas de Máxima Verossimilhança dos parâmetros dos modelos de regressão Normal-VEG e S_B de Johnson.

Parâmetro	Normal-VEG			S_B de Johnson		
	EMV	EP	P-valor	EMV	EP	P-valor
β_0	1,864	0,039	$0,00.10^{+00}$	1,905	0,040	$0,00.10^{+00}$
β_1	0,397	0,020	$0,00.10^{+00}$	0,391	0,020	$0,00.10^{+00}$
β_{2_1}	-0,329	0,038	$0,00.10^{+00}$	-0,360	0,040	$0,00.10^{+00}$
β_{2_2}	-0,780	0,041	$0,00.10^{+00}$	-0,800	0,044	$0,00.10^{+00}$
τ_0	3,474	0,616	$1,71.10^{-08}$	1,077	0,116	$0,00.10^{+00}$
τ_1	0,751	0,154	$1,11.10^{-06}$	0,342	0,098	$4,84.10^{-04}$
τ_{2_1}	0,424	0,163	$9,21.10^{-03}$	0,824	0,126	$5,46.10^{-11}$
τ_{2_2}	-0,608	0,291	$3,66.10^{-02}$	0,308	0,132	$1,96.10^{-02}$
λ	-1,190	0,319	$1,87.10^{-04}$			

Os dados também foram ajustados com o modelo de regressão GJS-t-Student com 4 graus de liberdade (LEMONTE; BAZÁN, 2016). E segundo os critérios de comparação de modelos apresentados na Tabela 24, observa-se que o modelo que ajusta melhor os dados é o modelo Normal-VEG-LL por apresentar menores valores dos critérios dentre os três modelos.

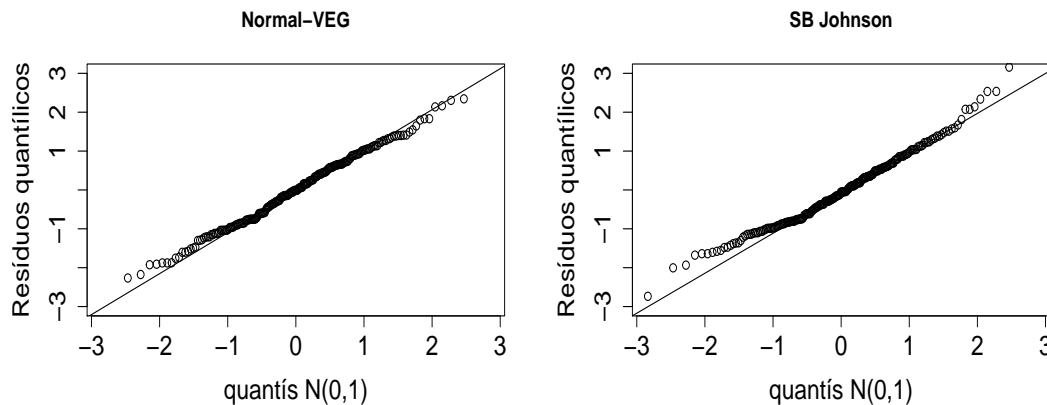
Depois, para testarmos a heterocedasticidade do modelo de regressão Normal-VEG-LL considerou-se o teste da razão de verossimilhança, onde testamos $H_0 : \tau_i = 0$ contra $H_1 : \tau_i \neq 0$, para $i = 1, 2, 3$. A estatística do teste de razão de verossimilhança resultou em 62,426 com 3 graus de liberdade e

Tabela 24 – Critério AIC e BIC para os modelos ajustados.

Modelo	Critério	
	AIC	BIC
Normal-VEG-LL	-997,219	-966,677
S_B de Johnson	-986,620	-959,471
GJS-t-Student(4)	-984,890	-957,741

com um p-valor = $1,781.10^{-13}$, que nos leva a rejeitar a hipótese nula e considerar o modelo heterocedástico para o ajuste de esses dados.

Por outro lado, para ver a adequabilidade do modelo, com as estimativas de máxima verossimilhança foi obtido o resíduo quantil aleatorizado (DUNN; SMYTH, 1996) para os modelos de regressão Normal-VEG-LL e modelo de regressão S_B de Johnson. A Figura 10 mostra o gráfico QQ-plot do resíduo quantil aleatorizado para ambos modelos, onde observa-se que esses resíduos se distribuem aproximadamente com a distribuição normal para ambos modelos indicando a adequabilidade dos modelos. Porém, da Figura 10, pode-se notar que o modelo normal-VEG se ajusta melhor a os dados. A normalidade dos resíduos é confirmado também pelo teste de normalidade de Kolmogorov-Smirnov (DANIEL *et al.*, 1978). A estatística de teste nesse caso resultou em 0,051 com um p-valor de 0,623 para o modelo Normal-VEG-LL, e em 0,055 com p-valor de 0,484 para o modelo S_B de Johnson indicando a normalidade dos resíduos para ambos modelos.

Figura 10 – QQ-plot de resíduos quantílicos aleatorizados para os modelos Normal-VEG e S_B Johnson

Desses resultados conclui-se que o modelo Normal-VEG-LL é adequado para modelar estes dados. Portanto, pode-se estimar a taxa de sobrevivência mediana e o respectivo intervalo de 95% de confiança dos pacientes com câncer colorretal para seis casos hipotéticos de pacientes A , B , C , D , E e F com características (covariáveis) dadas na Tabela 25.

Da Tabela 25 pode-se dizer, por exemplo, que para o caso B ou seja pacientes de sexo feminino e de origem hispânico tem uma taxa de sobrevivência da doença de 0,901, enquanto

Tabela 25 – Estimativa da mediana da taxa de sobrevivência e o intervalo de 95% de confiança para os seis pacientes hipotéticos com câncer colorretal.

Paciente	Sexo	Raça	Taxa de sobrevivência	
			EMV	95% Intervalo de confiança
<i>A</i>	Homem	Hispânico	0,856	(0,852 ; 0,861)
<i>B</i>	Mulher	Hispânico	0,901	(0,898 ; 0,904)
<i>C</i>	Homem	não-Hispânico branco	0,806	(0,804 ; 0,809)
<i>D</i>	Mulher	não-Hispânico branco	0,865	(0,864 ; 0,867)
<i>E</i>	Homem	não-Hispânico negro	0,713	(0,708 ; 0,718)
<i>F</i>	Mulher	não-Hispânico negro	0,797	(0,793 ; 0,800)

que para o caso *E* pacientes de sexo masculino e de origem não-hispânico negro têm uma taxa de sobrevivência de 0,713. Portanto pode-se dizer que as mulheres de origem hispânico que moram nos EUA tem a maior taxa de sobrevivência de câncer colorretal enquanto que os homens de origem não-hispânicos negro que moram nos EUA tem a menor taxa de sobrevivência de câncer colorretal. Concluindo com isso que as variáveis sexo e raça influenciam na taxa de sobrevivência do paciente com câncer colorretal nos EUA.

3.7 Conclusão

Neste Capítulo, foi desenvolvido o modelo de regressão respectivo para o modelo Normal-VEG fazendo uma reparametrização do modelo na mediana da variável resposta. O modelo de regressão foi construído como os modelos lineares generalizados, na qual a mediana e o parâmetro de dispersão estão relacionados a covariáveis conhecidas e parâmetros desconhecidos da regressão através de uma função de ligação. Inferências do modelo de regressão foram feitas pelo método de máxima verossimilhança. Também foram feitos estudos de simulação do modelo de regressão Normal-VEG homocedástico e heterocedástico para avaliar as propriedades dos EMV, onde verificou-se que eles possuem boas propriedades assintóticas. Um conjunto de dados de pacientes com câncer colorretal dos EUA foi ajustado com o modelo de regressão Normal-VEG e com o modelo de regressão S_B de Johnson. Observou-se que ambos modelos ajustam bem os dados, mas o melhor modelo dentre os dois conferido pela análise do resíduo quantil aleatorizado e os critérios de seleção de modelos é o modelo Normal-VEG. Portanto, o modelo proposto é adequado para fazer inferências da taxa de sobrevivência dos pacientes com câncer colorretal.

MODELO DE REGRESSÃO NORMAL-VEG: ENFOQUE BAYESIANO

No Capítulo 3, foi desenvolvido procedimentos inferências do modelo de regressão Normal-VEG sob uma perspectiva clássica, onde foi obtida os estimadores de máxima verossimilhança dos parâmetros do modelo. Entretanto, as inferências sobre os parâmetros é restrita à normalidade assintótica dos estimadores, e como muitas vezes não se têm um número grande de amostras a inferência dos parâmetros é limitada. Por outro lado, existe a abordagem Bayesiana que lida com este problema, e hoje em dia é utilizada por muitos pesquisadores. Por isso, neste capítulo será desenvolvido uma análise Bayesiana para o modelo de regressão Normal-VEG.

4.1 Modelo de regressão Normal-VEG

Como definido no Capítulo 3, considera-se novamente $\mathbf{Y} = (y_1, y_2, \dots, y_n)$ n realizações independentes da variável resposta Y , em que $Y_i \sim \text{Normal-VEG}(v_i, \delta_i, \lambda)$ para $i = 1, \dots, n$ com função densidade de probabilidade da forma

$$f_Y(y_i) = \frac{\delta_i \phi(\delta_i[Q(y_i; \lambda) - Q(v_i; \lambda)])}{y_i[-\log(y_i)]^{\lambda+1}}, \quad y_i \in (0, 1), \quad (4.1)$$

em que v_i e δ_i com a seguinte relação funcional

$$g_1(v_i) = \mathbf{x}_i^\top \boldsymbol{\beta},$$

$$g_2(\delta_i) = \mathbf{s}_i^\top \boldsymbol{\tau},$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ e $\boldsymbol{\tau} = (\tau_1, \dots, \tau_q)^\top$ são os vetores dos parâmetros desconhecidos da regressão e $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip})$ e $\mathbf{s}_i^\top = (s_{i1}, \dots, s_{iq})$ são as observações de p e q covariáveis conhecidas. Sejam $g_1 : (0, 1) \rightarrow \mathbb{R}$ e $g_2 : (0, \infty) \rightarrow \mathbb{R}$ funções de ligação para v e δ respectivamente.

Com esta relação, e dado uma amostra observada $\mathcal{D} = (\mathbf{y}, \mathbf{X}, \mathbf{S})$ pode-se escrever a função de verossimilhança de λ , $\boldsymbol{\beta}$ e $\boldsymbol{\tau}$ como

$$L(\lambda, \boldsymbol{\beta}, \boldsymbol{\tau} | \mathcal{D}) = \prod_{i=1}^n \delta_i \phi(\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)]) \left[\prod_{i=1}^n y_i \right]^{-1} \left[\prod_{i=1}^n [-\log(y_i)]^{\lambda+1} \right]^{-1} \quad (4.2)$$

4.2 Inferência Bayesiana

No contexto Bayesiano do modelo de regressão Normal-VEG, os parâmetros são considerados variáveis aleatórias e não mais constantes desconhecidas. Esta abordagem permite incorporar algum conhecimento a priori sobre os parâmetros, esse conhecimento pode ser modelado por uma função densidade de probabilidade *a priori* $\pi(\boldsymbol{\varphi})$, e com a informação dos dados contida na função de verossimilhança $L(\boldsymbol{\varphi} | \mathcal{D})$, pode-se obter através do teorema de Bayes a função densidade de probabilidade *a posteriori* $\pi(\boldsymbol{\varphi} | \mathcal{D})$ dos parâmetros, a qual pode-se expressar como:

$$\pi(\boldsymbol{\varphi} | \mathcal{D}) \propto L(\boldsymbol{\varphi} | \mathcal{D}) \pi(\boldsymbol{\varphi}),$$

em que $\boldsymbol{\varphi} = (\lambda, \boldsymbol{\beta}, \boldsymbol{\tau})$ é o vetor de parâmetros da regressão e \mathcal{D} é o vetor dos dados.

Neste trabalho é assumido que as distribuições *a priori* de λ , $\boldsymbol{\beta}$ e $\boldsymbol{\tau}$ são independentes, isto é

$$\pi(\lambda, \boldsymbol{\beta}, \boldsymbol{\tau}) = \pi(\lambda) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\tau}),$$

com $\lambda \sim N(0, \sigma_\lambda)$, $\boldsymbol{\beta} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_\beta)$, e $\boldsymbol{\tau} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_\tau)$, em que $N_r(\mathbf{0}, \boldsymbol{\Sigma}_0)$ denota a distribuição normal r -variada com vetor de médias $\mathbf{0}$ e matriz de covariâncias $\boldsymbol{\Sigma}_0$. Aqui todos os hiper-parâmetros são especificados para expressar prioris não-informativas. Assim com estas especificações das funções de densidade *a priori* e conjuntamente com a função de verossimilhança $L(\boldsymbol{\varphi} | \mathcal{D})$ dada como em (4.2), pode-se obter a função densidade *a posteriori* dos parâmetros como:

$$\pi(\boldsymbol{\varphi} | \mathcal{D}) \propto \prod_{i=1}^n \delta_i \phi(\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)]) \left\{ \prod_{i=1}^n [-\log(y_i)]^{\lambda+1} \right\}^{-1} \exp \left\{ -\frac{1}{2} \left(\frac{\lambda^2}{\sigma_\lambda} + \boldsymbol{\beta}' \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\beta} + \boldsymbol{\tau}' \boldsymbol{\Sigma}_\tau^{-1} \boldsymbol{\tau} \right) \right\} \quad (4.3)$$

Inferências dos parâmetros do modelo podem ser feitos calculando a função densidade de probabilidade *a posteriori*. Porém a densidade conjunta *a posteriori* de λ , $\boldsymbol{\beta}$ e $\boldsymbol{\tau}$ acima é analiticamente intratável. Por isso as inferências serão baseadas no método de simulação Monte Carlo via Cadeias de Markov (MCMC). Neste caso, é usado o algoritmo Metropolis Hasting. Para a implementação do algoritmo é seguido os seguintes passos:

- (1) Especifique um valor inicial de $\boldsymbol{\varphi}_{(0)}$, para o contador inicial $j = 0$,

- (2) gere um novo valor $\boldsymbol{\varphi}'$ de $r(\boldsymbol{\varphi}', \boldsymbol{\varphi}_j) = N_{p+q+1}(\boldsymbol{\varphi}_j, \tilde{\boldsymbol{\Sigma}})$, em que $\tilde{\boldsymbol{\Sigma}}$ é a matriz de covariâncias de $\boldsymbol{\varphi}$ e é a mesma em qualquer etapa,
- (3) atualize $\boldsymbol{\varphi}_{(j)}$ e faça $\boldsymbol{\varphi}_{(j+1)} = \boldsymbol{\varphi}'$ com probabilidade $p_j = \min\{1, \pi(\boldsymbol{\varphi}'|\mathcal{D})/\pi(\boldsymbol{\varphi}_{(j)}|\mathcal{D})\}$, ou mantenha $\boldsymbol{\varphi}_{(j)}$ com probabilidade $1 - p_j$,
- (4) repita o passo (2) e (3) incrementando o contador, até que o processo atinja a distribuição estacionária.

4.3 Critério de comparação de modelos

Na abordagem Bayesiana existem muitos métodos de comparação de modelos que selecionam aquele que melhor ajuste os dados. Um dos critérios mais utilizados é o *deviance information criterion* (DIC) proposto por Spiegelhalter *et al.* (2002), esse critério é baseado na média *a posteriori* do desvio $d(\boldsymbol{\varphi})$ que pode ser aproximada por

$$\bar{d} = \sum_{t=1}^T d(\boldsymbol{\varphi}_t)/T,$$

em que, t indica a t -ésima realização de um total de T realizações com

$$d(\boldsymbol{\varphi}) = -2 \sum_{i=1}^n \log [f_Y(y_i; \boldsymbol{\varphi})],$$

na qual $f_Y(\cdot; \cdot)$ é a f.d.p. do modelo dada como em (4.1). Usando amostras MCMC, o DIC também pode ser estimado por $\widehat{DIC} = \bar{d} + \widehat{\rho}_d = 2\bar{d} - \widehat{d}$, em que $\rho_d = E\{d(\boldsymbol{\varphi})\} - d\{E(\boldsymbol{\varphi})\}$ é o número de parâmetros efetivos, em que $d\{E(\boldsymbol{\varphi})\}$ é o desvio avaliado na média da posteriori e é estimado por

$$\widehat{D} = d\left(\frac{1}{T} \sum_{t=1}^T \boldsymbol{\lambda}^{(t)}, \frac{1}{T} \sum_{t=1}^T \boldsymbol{\beta}^{(t)}, \frac{1}{T} \sum_{t=1}^T \boldsymbol{\tau}^{(t)}\right).$$

O modelo que ajuste melhor os dados observados dentro de um conjunto de modelos será aquele que tem menor valor do DIC.

Outros critérios que também serão considerados neste trabalho são: o *Expected Akaike Information Criterion* (EAIC) (BROOKS *et al.*, 2002), e o *Expected Bayesian Information Criterion* (EBIC) (CARLIN; LOUIS, 2010), os quais são também baseados na média a posteriori do desvio, que pode ser aproximado por: $\widehat{EAIC} = \bar{d} + 2n(\boldsymbol{\varphi})$, e $\widehat{EBIC} = \bar{d} + n(\boldsymbol{\varphi})\log(n)$, em que $n(\boldsymbol{\varphi})$ é o número de parâmetros do modelo de regressão Normal-VEG. Similarmente ao critério DIC, valores menores de ambos critérios indicam melhor modelo.

Outro critério de comparação mais utilizado, considerado também neste trabalho é o critério da ordenada preditiva condicional (CPO) apresentadas por Gelfand, Dey e Chang (1992) e Geisser e Eddy (1979). Da qual denota-se \mathcal{D} sendo o conjunto de dados completos, e $\mathcal{D}^{(-i)}$

sendo os dados sem o i -ésimo elemento observado. A distribuição *a posteriori* de $\boldsymbol{\varphi}$ dado $\mathcal{D}^{(-i)}$ é denotado por $\pi(\boldsymbol{\varphi}|\mathcal{D}^{(-i)})$, $i = 1, \dots, n$. Pode-se obter o CPO_i para a i -ésima observação como:

$$CPO_i = \int_{\boldsymbol{\varphi} \in \Phi} f_Y(y_i; \boldsymbol{\varphi}) \pi(\boldsymbol{\varphi}|\mathcal{D}^{(-i)}) d\boldsymbol{\varphi} = \left\{ \int_{\boldsymbol{\varphi}} \frac{\pi(\boldsymbol{\varphi}|\mathcal{D})}{f_Y(y_i; \boldsymbol{\varphi})} d\boldsymbol{\varphi} \right\}^{-1}. \quad (4.4)$$

Para o modelo proposto, não é possível encontrar uma expressão fechada de CPO_i . Porém uma estimativa Monte Carlo de CPO_i pode ser obtido usando amostras MCMC da distribuição *a posteriori* $\pi(\boldsymbol{\varphi}|\mathcal{D})$. Para um tamanho de amostra T obtida de $\pi(\boldsymbol{\varphi}|\mathcal{D})$, uma aproximação Monte Carlo de CPO_i é dada por

$$\widehat{CPO}_i = \left\{ \frac{1}{T} \sum_{t=1}^T \frac{1}{f_Y(y_i; \boldsymbol{\varphi}^{(t)})} \right\}^{-1}.$$

Para comparar modelos é utilizada a estatística *log pseudo marginal likelihood* (LPML) (CANCHO; ORTEGA; PAULA, 2010) definido por $LPML = \sum_{i=1}^n \log(\widehat{CPO}_i)$. Valores maiores de LPML selecionam o melhor modelo.

4.4 Análise de diagnóstico bayesiano

Em análise de regressão após o ajuste do modelo é importante verificar as suposições do modelo com o propósito de obter um modelo mais robusto, encontrando pontos que possam influenciar nos resultados finais da análise. Cook (1986) diz que o melhor modelo é aquele que sob pequenas mudanças ele fica relativamente estável. Um dos esquemas de perturbação mais conhecidos é o de eliminação de casos (COOK; WEISBERG, 1982) na qual os efeitos são estudados removendo completamente os casos da análise. Essa ideia é essencial para construir nossa metodologia global de influência bayesiana com a finalidade de identificar a observação que faz uma grande influência na análise do modelo.

Seja $D_{\psi}(P, P_{(-i)})$ a ψ -divergência entre P e $P_{(-i)}$, em que P é a distribuição *a posteriori* de $\boldsymbol{\varphi}$ para dados completos e $P_{(-i)}$ é a distribuição *a posteriori* de $\boldsymbol{\varphi}$ sem o i -ésimo caso, especificamente,

$$D_{\psi}(P, P_{(-i)}) = \int \psi \left(\frac{\pi(\boldsymbol{\varphi}|\mathcal{D}^{(-i)})}{\pi(\boldsymbol{\varphi}|\mathcal{D})} \right) \pi(\boldsymbol{\varphi}|\mathcal{D}) d\boldsymbol{\varphi},$$

em que ψ é uma função convexa com $\psi(1) = 0$, diferentes escolhas de ψ são dadas em Dey e Birmiwal (1994) e Pardo (2006). Por exemplo se $\psi(z) = -\log(z)$ obtem-se a divergência Kullback-Leibler (divergência K-L), se $\psi(z) = (z-1)\log(z)$ tem-se a J -distância (ou versão simétrica da divergência K-L), se $\psi(z) = 0.5|z-1|$ tem-se a distancia variacional ou a norma L_1 ; e, se $\psi(z) = z(1/z-1)^2$ tem-se a divergência qui-quadrado (divergência χ^2)

A relação entre o CPO e a ψ -divergência é dado por:

$$D_{\psi}(P, P_{(-i)}) = E_{\boldsymbol{\varphi}|\mathcal{D}} \left[\psi \left(\frac{CPO_i}{f_Y(y_i; \boldsymbol{\varphi})} \right) \right], \quad (4.5)$$

em que o valor esperado é tomado com respeito a distribuição conjunta *a posteriori* $\pi(\boldsymbol{\varphi}|\mathcal{D})$. Como caso particular tem-se que a divergência K-L pode ser expressada como

$$\begin{aligned} D_{\text{K-L}}(P, P_{(-i)}) &= -E_{\boldsymbol{\varphi}|\mathcal{D}} \{\log(\text{CPO}_i)\} + E_{\boldsymbol{\varphi}|\mathcal{D}} \{\log [f_Y(y_i; \boldsymbol{\varphi})]\} \\ &= -\log(\text{CPO}_i) + E_{\boldsymbol{\varphi}|\mathcal{D}} \{\log [f_Y(y_i; \boldsymbol{\varphi})]\}. \end{aligned}$$

De (4.5) pode-se calcular $D_{\psi}(P, P_{(-i)})$ por amostragem da distribuição *a posteriori* de $\boldsymbol{\varphi}$ por médio do método MCMC.

Seja $\boldsymbol{\varphi}^{(1)}, \dots, \boldsymbol{\varphi}^{(T)}$ amostras de tamanho T de $\pi(\boldsymbol{\varphi}|\mathcal{D})$. Então a estimativa de Monte Carlo de $D_{\psi}(P, P_{(-i)})$ é dada por

$$\widehat{D}_{\psi}(P, P_{(-i)}) = \frac{1}{T} \sum_{t=1}^T \psi \left(\frac{\widehat{\text{CPO}}_i}{f_Y(y_i; \boldsymbol{\varphi}^{(t)})} \right). \quad (4.6)$$

Como caso particular tem-se a estimativa Monte Carlo para a divergência K-L $D_{\text{K-L}}(P, P_{(-i)})$, que é dada por

$$\widehat{D}_{\text{K-L}}(P, P_{(-i)}) = -\log(\widehat{\text{CPO}}_i) + \frac{1}{T} \sum_{t=1}^T \log [f_Y(y_i; \boldsymbol{\varphi}^{(t)})].$$

$D_{\psi}(P, P_{(-i)})$ mede o efeito de eliminar o i -ésimo caso dos dados completos na distribuição conjunta *a posteriori* de $\boldsymbol{\varphi}$. Peng e Dey (1995) e Weiss (1996) afirmam que pode ser difícil obter o ponto de corte da medida de divergência, de modo que se possa determinar se um pequeno conjunto de observações dos dados completos é influente ou não. E por isso nesse contexto é usado a proposta de calibração dada por Peng e Dey (1995) e Weiss (1996). Assim considere uma moeda viesada com probabilidade de sucesso p . Assim, a medida ψ -divergência entre a moeda viesada e não viesada é

$$D_{\psi}(f_0, f_1) = \int \psi \left(\frac{f_0(x)}{f_1(x)} \right) f_1(x) dx,$$

em que $f_0(x) = p^x(1-p)^{1-x}$ e $f_1(x) = 0.5$, $x = 0; 1$. E se $D_{\psi}(f_0, f_1) = d_{\psi}(p)$, pode-se verificar facilmente que d_{ψ} satisfaz a seguinte equação:

$$d_{\psi}(p) = \frac{\psi(2p) + \psi(2(1-p))}{2}.$$

Não é complicado notar que d_{ψ} aumenta à medida que p se afasta de 0,5. Além disso, $d_{\psi}(p)$ é simétrica em torno de $p = 0,5$ e também d_{ψ} atinge seu valor mínimo quando $p = 0,5$. Neste ponto $d_{\psi}(0,5) = 0$ e $f_0 = f_1$. Além disso se se considera $p > 0,80$ (ou $p < 0,20$) como uma moeda de forte viés, então $d_{L_1}(0,80) = 0,30$. Esta equação implica que o i -ésimo caso pode ser considerado como influente quando $d_{L_1} > 0,30$. Assim também se é usado a divergência K-L, pode-se considerar uma observação influente quando $d_{\text{K-L}} > 0,223$. Similarmente se é usada a J -distância, uma observação é considerada influente se $d_J > 0,416$. E se é usada a divergência χ^2 , uma observação é influente quando $d_{\chi^2} > 0,562$.

4.5 Estudo de simulação

Nesta seção é apresentado um estudo de simulação com o objetivo de avaliar o comportamento das estimativas bayesianas dos parâmetros do modelo de regressão Normal-VEG sob a perda quadrática e perda absoluta.

Foram geradas amostras de tamanhos $n = 50, 100, 200$ e 400 do modelo de regressão Normal-VEG(λ, v_i, δ_i) da mesma forma que na Seção 3.5, com parâmetros fixados em $\beta_0 = -3, \beta_1 = -2, \tau_0 = 1, \tau_1 = 1$, mas para o parâmetro λ foram considerados os valores $-0,4, 0,4, 0,1$ e $-0,1$. Para executar o algoritmo Metropolis-Hasting, foram consideradas as seguintes priors independentes: $\lambda \sim N(0, 1)$; $\beta \sim N_2(\mathbf{0}, \Sigma_0)$ e $\tau \sim N_2(\mathbf{0}, \Sigma_0)$ com $\Sigma_0 = \text{diag}\{100\}$.

A finalidade deste estudo de simulação é mostrar o comportamento das estimativas bayesianas com base nas estatísticas frequentistas como são: a média, o desvio padrão (DP), o viés, a raiz quadrada do erro quadrático médio (REQM) e a probabilidade de cobertura (PC).

Portanto para cada conjunto de dados simulados, foram obtidos as estatísticas de resumo e os intervalos HPD de 95% de credibilidade dos parâmetros do modelo. Considerou-se 25000 amostras MCMC da posteriori, geradas para cada parâmetro, das quais as primeiras 5000 iterações foram desconsideradas para eliminar o efeito dos valores iniciais e assim obter uma amostra de tamanho 20000. A autocorrelação destes valores amostrados é reduzida tomando um espaçamento de tamanho 10, resultando em uma amostra final de tamanho 2000. A convergência das cadeias foi monitorada usando os métodos de Cowles e Carlin (1996). Para cada configuração foram feitas 1000 replicas, e das estimativas de cada parâmetro foram obtidas a média, o desvio padrão (DP), a raiz do erro quadrático médio (REQM) e a probabilidade de cobertura (PC). Os resultados obtidos são apresentados nas Tabelas 26, 27, 28 e 29, das quais temos que:

- Na Tabela 26 observa-se os resultados para o modelo de regressão Normal-VEG com parâmetro $\lambda = -0,4, \beta_0 = -3, \beta_1 = -2, \tau_0 = 1, \tau_1 = 1$. Note que as estimativas bayesianas dos parâmetros do modelo se aproximam em média do verdadeiro valor dos parâmetros. A raiz do erro quadrático médio diminui, o valor da probabilidade de cobertura tornando-se ao redor do nível nominal de 95% como esperado. Tudo isto a medida que o tamanho da amostra aumenta para ambos casos de perdas, mas note que com a perda quadrática obtém-se uma melhor estimativa para λ .
- Na Tabela 27 observa-se resultados desta vez quando $\lambda = 0,4$, e os outros parâmetros mantendo-se fixos como no caso anterior. Pode-se notar que os resultados são parecidos ao caso anterior, isto é as estimativas dos parâmetros perto dos verdadeiros valores em media. O viés, a raiz do erro quadrático médio diminuindo e tendendo para zero, a probabilidade de cobertura ao redor do nível nominal de 95%. Tudo isto à medida que o tamanho da amostra aumenta, para ambos casos, mas cabe notar que os valores do viés com a perda absoluta é menor do que com a perda quadrática, então obtém-se melhores estimativas

para os parâmetros usando a perda absoluta.

- Na Tabela 28 tem-se os resultados da simulação do modelo desta vez com parâmetro $\lambda = -0,1$, mantendo os demais parâmetros fixos, e observa-se que em termos gerais os resultados são os mesmos que o caso anterior, mas obtendo-se melhores resultados para um tamanho de amostra 400 e com perda absoluta.
- Na Tabela 29 tem-se os resultados da simulação do modelo com o parâmetro $\lambda = 0,1$, resultados similares em termos gerais que os outros casos, e pode-se observar que a REQM vai diminuindo a medida que o tamanho da amostra aumenta, e é menor com a perda absoluta.

Concluindo que as estimativas Bayesianas dos parâmetros do modelo de regressão Normal-VEG reproduzem os verdadeiros valores dos parâmetros quando o tamanho da amostra aumenta.

Tabela 26 – Estimativas bayesianas sob perda quadrática e perda absoluta dos parâmetros do modelo de regressão Normal-VEG, desvio padrão (DP), viés, a raiz do erro quadrático médio (REQM) e a probabilidade de cobertura (PC), baseados em 1000 simulações do modelo com parâmetros $\lambda = -0,4$, $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$.

n	Parâmetro	Perda Quadrática				Perda Absoluta				PC
		Média	DP	Viés	REQM	Média	DP	Viés	REQM	
50	β_0	-3,008	0,190	-0,008	0,190	-3,007	0,190	-0,007	0,190	0,935
	β_1	-1,990	0,272	0,010	0,272	-1,991	0,272	0,009	0,272	0,945
	τ_0	1,032	0,953	0,032	0,953	1,060	0,930	0,060	0,931	0,980
	τ_1	1,030	0,558	0,030	0,558	1,024	0,552	0,024	0,553	0,965
	λ	-0,368	0,820	0,032	0,821	-0,358	0,798	0,042	0,799	0,983
100	β_0	-2,996	0,131	0,004	0,131	-2,995	0,132	0,005	0,132	0,940
	β_1	-2,005	0,194	-0,005	0,194	-2,005	0,194	-0,005	0,194	0,936
	τ_0	1,050	0,672	0,050	0,673	1,065	0,664	0,065	0,667	0,953
	τ_1	1,015	0,391	0,015	0,391	1,012	0,388	0,012	0,388	0,955
	λ	-0,360	0,575	0,040	0,576	-0,355	0,567	0,045	0,569	0,969
200	β_0	-3,002	0,090	-0,002	0,090	-3,002	0,090	-0,002	0,090	0,944
	β_1	-1,996	0,131	0,004	0,131	-1,997	0,131	0,003	0,131	0,941
	τ_0	1,047	0,450	0,047	0,452	1,054	0,448	0,054	0,451	0,966
	τ_1	1,022	0,263	0,022	0,264	1,020	0,263	0,020	0,263	0,960
	λ	-0,355	0,388	0,045	0,390	-0,353	0,386	0,047	0,389	0,963
400	β_0	-3,004	0,067	-0,004	0,067	-3,003	0,067	-0,003	0,067	0,930
	β_1	-1,995	0,095	0,005	0,095	-1,995	0,095	0,005	0,095	0,943
	τ_0	1,008	0,337	0,008	0,337	1,011	0,336	0,011	0,336	0,938
	τ_1	1,000	0,189	0,000	0,189	0,999	0,189	-0,001	0,189	0,945
	λ	-0,394	0,289	0,006	0,289	-0,393	0,288	0,007	0,288	0,939

Tabela 27 – Estimativas bayesianas sob perda quadrática e perda absoluta dos parâmetros do modelo de regressão Normal-VEG, DP, viés, REQM e PC, baseados em 1000 simulações do modelo com parâmetros $\lambda = 0,4$, $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$.

n	Parâmetro	Perda Quadrática				Perda Absoluta				
		Média	DP	Viés	REQM	Média	DP	Viés	REQM	PC
50	β_0	-3,140	0,545	-0,140	0,562	-3,107	0,542	-0,107	0,552	0,938
	β_1	-1,899	0,826	0,101	0,832	-1,915	0,822	0,085	0,826	0,947
	τ_0	0,942	0,417	-0,058	0,421	0,953	0,414	-0,047	0,416	0,959
	τ_1	0,966	0,408	-0,034	0,409	0,964	0,407	-0,036	0,409	0,952
	λ	0,345	0,292	-0,055	0,297	0,341	0,289	-0,059	0,295	0,961
100	β_0	-3,057	0,366	-0,057	0,370	-3,040	0,365	-0,040	0,367	0,943
	β_1	-1,969	0,583	0,031	0,583	-1,977	0,582	0,023	0,582	0,939
	τ_0	0,969	0,292	-0,031	0,294	0,974	0,291	-0,026	0,292	0,945
	τ_1	0,981	0,283	-0,019	0,283	0,981	0,282	-0,019	0,283	0,932
	λ	0,370	0,206	-0,030	0,208	0,368	0,205	-0,032	0,207	0,954
200	β_0	-3,031	0,255	-0,031	0,256	-3,023	0,255	-0,023	0,255	0,944
	β_1	-1,986	0,390	0,014	0,390	-1,990	0,390	0,010	0,390	0,949
	τ_0	0,980	0,187	-0,020	0,188	0,983	0,187	-0,017	0,187	0,965
	τ_1	0,987	0,192	-0,013	0,192	0,987	0,191	-0,013	0,192	0,950
	λ	0,383	0,133	-0,017	0,134	0,382	0,133	-0,018	0,134	0,953
400	β_0	-3,022	0,177	-0,022	0,178	-3,018	0,177	-0,018	0,178	0,949
	β_1	-1,983	0,271	0,017	0,271	-1,985	0,271	0,015	0,271	0,955
	τ_0	0,981	0,142	-0,019	0,143	0,983	0,142	-0,017	0,143	0,939
	τ_1	1,003	0,129	0,003	0,129	1,003	0,129	0,003	0,129	0,952
	λ	0,388	0,101	-0,012	0,101	0,388	0,101	-0,012	0,101	0,941

Tabela 28 – Estimativas bayesianas sob perda quadrática e perda absoluta dos parâmetros do modelo de regressão Normal-VEG, DP, viés, REQM e PC, baseados em 1000 simulações do modelo com parâmetros $\lambda = -0,1$, $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$.

n	Parâmetro	Perda Quadrática				Perda Absoluta				
		Média	DP	Viés	REQM	Média	DP	Viés	REQM	PC
50	β_0	-3,027	0,278	-0,027	0,279	-3,021	0,279	-0,021	0,279	0,941
	β_1	-1,974	0,408	0,026	0,408	-1,977	0,408	0,023	0,408	0,949
	τ_0	0,981	0,670	-0,019	0,670	1,000	0,658	0,000	0,658	0,976
	τ_1	1,007	0,470	0,007	0,470	1,002	0,468	0,002	0,467	0,961
	λ	-0,112	0,558	-0,012	0,558	-0,110	0,547	-0,010	0,547	0,977
100	β_0	-3,003	0,192	-0,003	0,192	-3,000	0,192	0,000	0,192	0,940
	β_1	-2,001	0,291	-0,001	0,291	-2,002	0,291	-0,002	0,291	0,933
	τ_0	1,009	0,480	0,009	0,480	1,019	0,477	0,019	0,477	0,951
	τ_1	0,996	0,330	-0,004	0,330	0,994	0,329	-0,006	0,329	0,955
	λ	-0,096	0,397	0,004	0,397	-0,095	0,394	0,005	0,394	0,957
200	β_0	-3,008	0,132	-0,008	0,132	-3,007	0,133	-0,007	0,133	0,948
	β_1	-1,991	0,198	0,009	0,198	-1,991	0,198	0,009	0,198	0,946
	τ_0	1,015	0,320	0,015	0,320	1,019	0,319	0,019	0,319	0,965
	τ_1	1,008	0,223	0,008	0,223	1,006	0,223	0,006	0,223	0,950
	λ	-0,084	0,268	0,016	0,268	-0,084	0,267	0,016	0,267	0,961
400	β_0	-3,008	0,097	-0,008	0,098	-3,007	0,097	-0,007	0,098	0,933
	β_1	-1,990	0,142	0,010	0,142	-1,991	0,142	0,009	0,142	0,940
	τ_0	0,997	0,239	-0,003	0,239	0,999	0,239	-0,001	0,239	0,937
	τ_1	0,995	0,158	-0,005	0,158	0,995	0,158	-0,005	0,158	0,943
	λ	-0,104	0,199	-0,004	0,199	-0,103	0,199	-0,003	0,198	0,944

Tabela 29 – Estimativas bayesianas sob perda quadrática e perda absoluta dos parâmetros do modelo de regressão Normal-VEG, DP, viés, REQM e PC, baseados em 1000 simulações do modelo com parâmetros $\lambda = 0, 1$, $\beta_0 = -3$, $\beta_1 = -2$, $\tau_0 = 1$, $\tau_1 = 1$.

n	Parâmetro	Perda Quadrática				Perda Absoluta				
		Média	DP	Viés	REQM	Média	DP	Viés	REQM	PC
50	β_0	-3,055	0,361	-0,055	0,365	-3,043	0,362	-0,043	0,364	0,939
	β_1	-1,954	0,538	0,046	0,540	-1,959	0,538	0,041	0,539	0,953
	τ_0	0,953	0,541	-0,047	0,543	0,968	0,535	-0,032	0,535	0,972
	τ_1	0,995	0,437	-0,005	0,436	0,991	0,436	-0,009	0,436	0,958
	λ	0,063	0,434	-0,037	0,436	0,062	0,428	-0,038	0,429	0,979
100	β_0	-3,014	0,248	-0,014	0,248	-3,008	0,248	-0,008	0,248	0,942
	β_1	-1,994	0,383	0,006	0,383	-1,997	0,383	0,003	0,383	0,936
	τ_0	0,989	0,390	-0,011	0,390	0,997	0,388	-0,003	0,388	0,946
	τ_1	0,988	0,306	-0,012	0,306	0,986	0,306	-0,014	0,306	0,946
	λ	0,086	0,310	-0,014	0,310	0,086	0,308	-0,014	0,308	0,961
200	β_0	-3,017	0,170	-0,017	0,171	-3,014	0,170	-0,014	0,171	0,946
	β_1	-1,983	0,260	0,017	0,260	-1,985	0,260	0,015	0,260	0,945
	τ_0	1,000	0,259	0,000	0,258	1,004	0,258	0,004	0,258	0,964
	τ_1	1,002	0,206	0,002	0,206	1,001	0,206	0,001	0,206	0,948
	λ	0,104	0,208	0,004	0,208	0,104	0,208	0,004	0,208	0,964
400	β_0	-3,013	0,125	-0,013	0,126	-3,011	0,126	-0,011	0,126	0,934
	β_1	-1,985	0,186	0,015	0,187	-1,986	0,187	0,014	0,187	0,943
	τ_0	0,992	0,193	-0,008	0,193	0,994	0,192	-0,006	0,192	0,934
	τ_1	0,993	0,145	-0,007	0,145	0,993	0,145	-0,007	0,145	0,941
	λ	0,092	0,154	-0,008	0,154	0,092	0,154	-0,008	0,154	0,941

4.6 Aplicação

Nesta seção é apresentado uma aplicação a dados reais do modelo de regressão Normal-VEG sob o enfoque bayesiano, para o qual considerou-se o mesmo conjunto de dados descrito na aplicação do Capítulo 3, onde tinha-se dados de pacientes com câncer colorretal dos anos 2006 a 2010 nos EUA, da qual é considerada como variável resposta a taxa de sobrevivência dos pacientes. O objetivo segue sendo avaliar se existem diferenças significativas na taxa de sobrevivência de câncer colorretal entre homens e mulheres, assim como também entre pacientes de diferentes raças. Assim tem-se novamente que:

- y_i : taxa de sobrevivência, para $i = 1, \dots, 220$;
- x_{1i} : sexo (homem=0, mulher=1);
- x_{2i} : raça (não hispânico branco, não hispânico negro, hispânico).

Os dados foram ajustados com o modelo de regressão descritos nas Seções 4.1 e 4.2, ou seja é modelado a mediana (v) é o parâmetro de dispersão (δ) com todas as covariáveis. Como no caso clássico considerou-se as seguintes funções de ligação: logística, probit, cauchy, complemento

log-log e log-log para a mediana e a função de ligação logarítmica para o parâmetro de dispersão, isto é:

$$\begin{aligned} g_1(v_i) &= \beta_0 + \beta_1 x_{1i} + \beta_{2_1} x_{2_1i} + \beta_{2_2} x_{2_2i}, \text{ e} \\ g_2(\delta_i) &= \log(\delta_i) = \tau_0 + \tau_1 x_{1i} + \tau_{2_1} x_{2_1i} + \tau_{2_2} x_{2_2i}, \quad i = 1, \dots, 220, \end{aligned} \quad (4.7)$$

em que

$$x_{1i} = \begin{cases} 1, & \text{se mulher;} \\ 0, & \text{se homem} \end{cases},$$

$$x_{2_1i} = \begin{cases} 1, & \text{se não-hispânico branco;} \\ 0, & \text{outro caso} \end{cases} \quad \text{e} \quad x_{2_2i} = \begin{cases} 1, & \text{se não-hispânico negro;} \\ 0, & \text{outro caso.} \end{cases}$$

Considerou-se prioris pouco informativas para os parâmetros, isto é: $\lambda \sim N(0, 1)$, $\boldsymbol{\beta} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_\beta)$, e $\boldsymbol{\tau} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_\tau)$ com $\boldsymbol{\Sigma}_0 = \text{diagonal}\{100\}$. Para obter a distribuição a posteriori dos parâmetros foi utilizado o método MCMC, na qual os algoritmos MCMC foram implementados na linguagem R (R Core Team, 2015). Foram usadas 25.000 iterações MCMC em todos os cálculos das quais foram excluídas 5.000 iterações para eliminar o efeito dos valores iniciais, e para reduzir a autocorrelação entre os valores amostrados foi tomado um espaçamento de tamanho 5, e assim resultados a posteriori são baseados em 4.000 iterações da cadeia de Markov. A convergência das cadeias foi monitorada usando os métodos de Cowles e Carlin (1996).

Com as amostras a posteriori, foi obtida as estimativas Monte Carlo dos critérios DIC, EAIC, EBIC e LPML mencionados na Seção 4.3 cujos valores para cada modelo estão dados na Tabela 30, da qual de acordo com todos os critérios o modelo que melhor ajusta os dados, como no enfoque clássico, é o modelo Normal-VEG com a função de ligação log-log o qual é denominado de modelo de regressão Normal-VEG-LL.

Tabela 30 – Critérios de comparação de modelos DIC, EAIC, EBIC, e LPML do modelo de regressão Normal-VEG com diferentes funções de ligações para os dados de câncer colorretal

Função de ligação	Critério			
	DIC	EAIC	EBIC	LPML
Logístico	-995,771	-986,688	-956,145	495,769
Probit	-994,348	-984,993	-954,450	496,388
Cauchy	-993,009	-984,211	-953,669	495,222
Com-log-log	-990,889	-981,592	-951,049	494,635
Log-log	-996,574	-987,735	-957,192	497,308

Os gráficos de traço para os parâmetros do modelo de regressão Normal-VEG-LL são mostrados na Figura 11, o qual indica a convergência das cadeias.

Na Tabela 31, são apresentadas as estimativas bayesianas sob a perda quadrática e perda absoluta dos parâmetros do modelo de regressão Normal-VEG-LL e do modelo de regressão S_B de Johnson, assim como também o intervalo HPD (*highest posterior density*) de 95% de

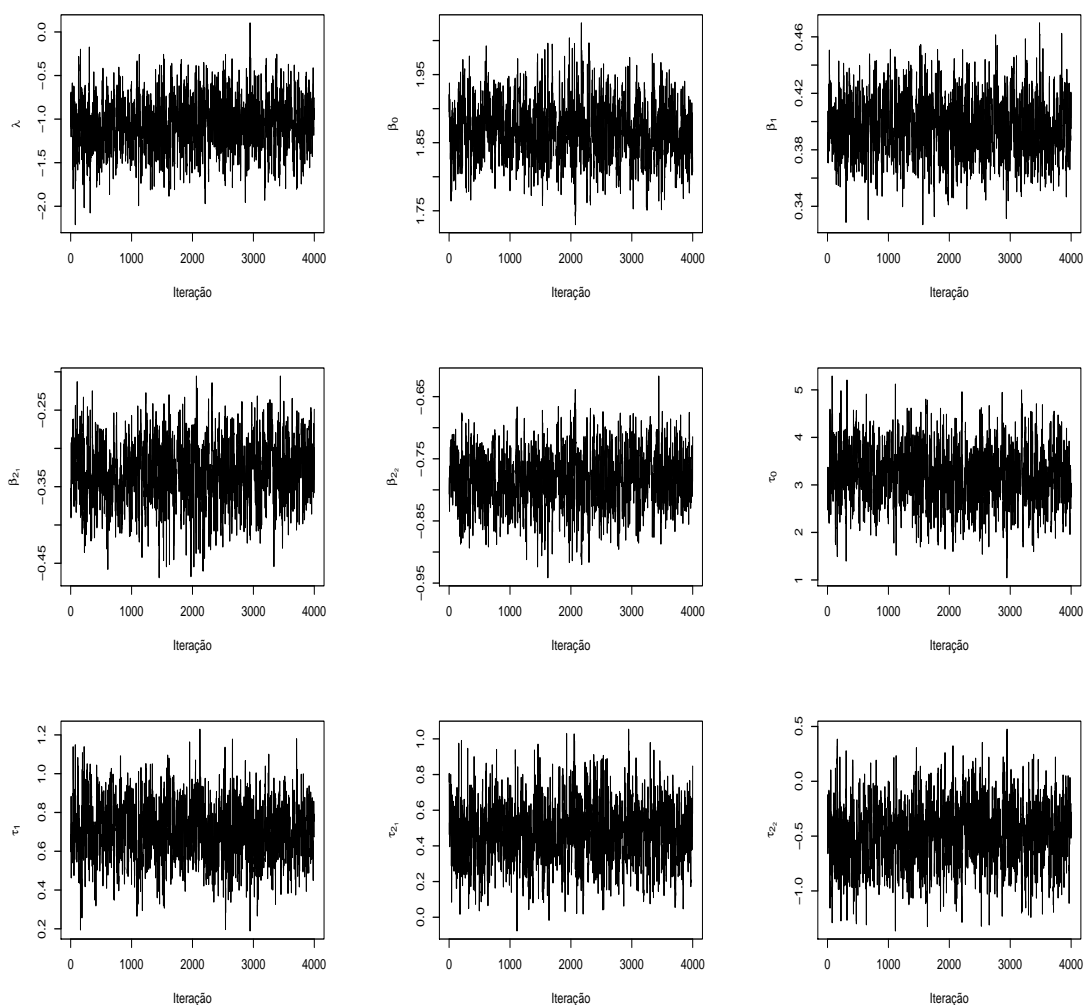


Figura 11 – Gráfico de traço para os parâmetros do modelo Normal-VEG-LL para os dados de câncer colorretal

credibilidade. O efeito das covariáveis na taxa de sobrevivência mediana são estatisticamente significativas com um nível de significância de 5% para ambos modelos. Além disso, para comparar modelos, os dados também foram ajustados com o modelo de regressão GJS-t-Student introduzida por [Lemonte e Bazán \(2016\)](#) fixado com 4 graus de liberdade e função de ligação log-log. Para comparar os modelos foi usada as estimativas Monte Carlo dos critérios DIC, EAIC, EBIC e LPML dos modelos Normal-VEG, S_B de Johnson e GJS-t-Student(4) todos com função de ligação log-log. Os valores dos critérios são apresentados na Tabela 32, a qual de acordo a todos os critérios o modelo GJS-t-Student(4) é melhor que o modelo S_B de Johnson, mas não é melhor que o modelo Normal-VEG-LL.

A Figura 12 (lado esquerdo), mostra o gráfico da densidade marginal a posteriori para o parâmetro λ do modelo de regressão Normal-VEG-LL, da qual pode-se dizer que a distribuição a posteriori de λ é simétrica. Também da Tabela 31 tem-se que a media a posteriori de λ é -1,061 com um intervalo HPD de (-1,617 ; -0,467) de 95% de credibilidade.

Tabela 31 – Estimativas e intervalo HPD (95%) dos parâmetros do modelo de regressão Normal-VEG e S_B de Johnson com função de ligação log-log para os dados de câncer colorretal

Parâmetro	Normal-VEG			S_B de Johnson		
	Média	Mediana	Intervalo HPD (95%)	Média	Mediana	Intervalo HPD (95%)
β_0	1,869	1,868	(1,792 ; 1,952)	1,902	1,902	(1,823 ; 1,981)
β_1	0,396	0,396	(0,361 ; 0,437)	0,392	0,392	(0,353 ; 0,431)
β_{2_1}	-0,332	-0,330	(-0,414 ; -0,260)	-0,358	-0,358	(-0,437 ; -0,281)
β_{2_2}	-0,783	-0,782	(-0,870 ; -0,703)	-0,798	-0,797	(-0,882 ; -0,712)
τ_0	3,202	3,209	(1,994 ; 4,228)	1,053	1,054	(0,811 ; 1,279)
τ_1	0,703	0,703	(0,426 ; 0,991)	0,341	0,344	(0,155 ; 0,534)
τ_{2_1}	0,476	0,472	(0,135 ; 0,778)	0,831	0,826	(0,593 ; 1,086)
τ_{2_2}	-0,490	-0,497	(-1,030 ; 0,026)	0,312	0,313	(0,047 ; 0,586)
λ	-1,061	-1,066	(-1,617 ; -0,467)			

Tabela 32 – Critério DIC, EAIC, EBIC e LPML para os modelos de regressão Normal-VEG, S_B de Johnson, e GJS-t-Student(4) com função de ligação log-log para os dados de câncer colorretal

Modelo	Critério			
	DIC	EAIC	EBIC	LPML
Normal-VEG-LL	-996,574	-987,735	-957,192	497,308
S_B de Johnson	-987,162	-978,800	-951,651	492,121
GJS-t-Student(4)	-990,811	-983,005	-955,966	495,770

Por outro lado para avaliar a adequação do modelo de regressão Normal-VEG-LL, foram obtidos os resíduos quantílicos aleatorizados normalizados a posteriori (DUNN; SMYTH, 1996). O gráfico QQ-plot dos resíduos pode-se observar na Figura 12 (lado direito), o qual indica que o modelo Normal-VEG-LL produz um ajuste razoável para os dados de câncer colorretal. Estes resíduos também foram obtidos para o modelo de regressão S_B de Johnson, onde o QQ-plot pode-se observar na Figura 12 (lado central), da qual pode-se notar que não faz melhor ajuste que o modelo Normal-VEG.

Também, com o objetivo de identificar possíveis observações influentes que possam afetar a análise da regressão, foi calculado as estimativas Monte Carlo das medidas de ψ -divergência para o modelo de regressão Normal-VEG-LL. A Figura 13 mostra os gráficos das medidas de divergência K-L, J, L_1 , e χ^2 , das quais observa-se que as observações 39, 54 e 122 podem ser possíveis observações influentes na distribuição a posteriori. As medidas de ψ -divergência detectaram as observações 39, 54 e 122 como casos potencialmente influentes, para o qual com a finalidade de revelar o impacto dessas observações na distribuição a posteriori dos parâmetros do modelo de regressão Normal-VEG-LL, foi reajustado o modelo para os seguintes casos: primeiro, foi eliminado individualmente cada um desses três casos do conjunto de dados originais, depois foram eliminados dois a dois, e por último os três juntos. A Tabela 35 apresenta as mudanças relativas (MRs) nas estimativas após retirar as observações influentes individualmente, dois a dois e por último as três de uma só vez, também são apresentados os intervalos HPD de 95% de credibilidade para as novas estimativas dos parâmetros obtidas ao retirar as observações. As mudanças relativas para cada parâmetro estimados é definido por

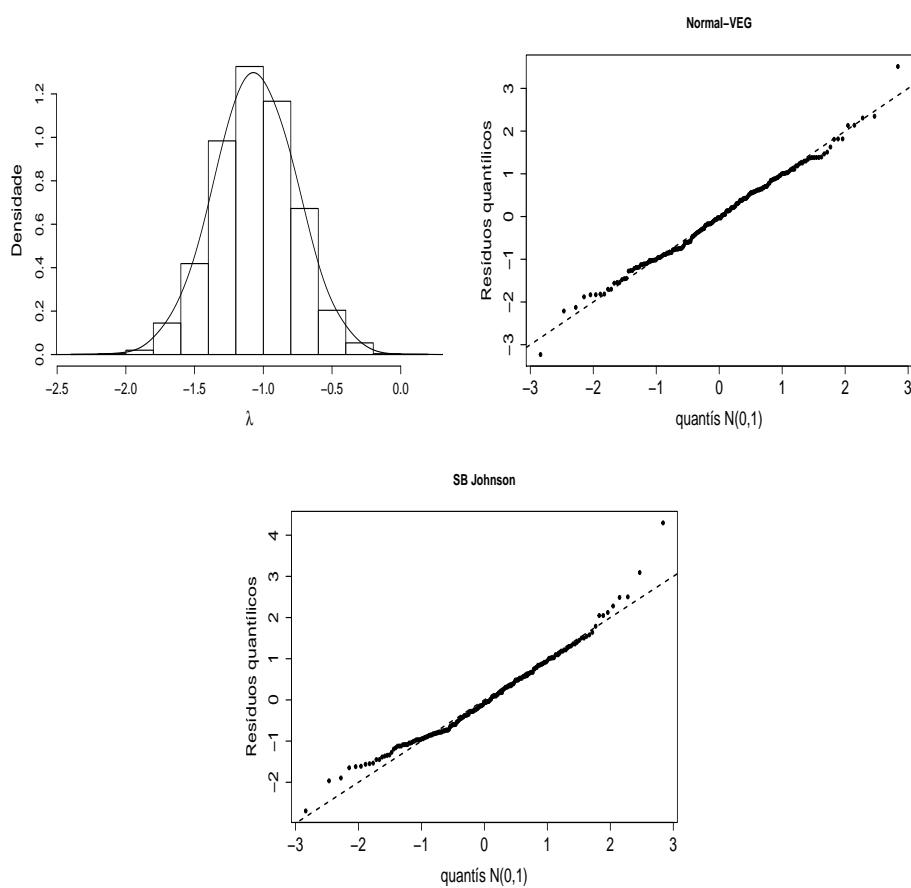


Figura 12 – Densidade marginal a posteriori do parâmetro λ do modelo de regressão Normal-VEG-LL (lado esquerdo), o QQ-plot dos resíduos quantílicos aleatorizados normalizados a posteriori do modelo Normal-VEG-LL (lado direito) e do modelo de regressão S_B de Johnson com ligação log-log (lado central)

$MR_{\varphi_j} = (\hat{\varphi}_j - \hat{\varphi}_{j(I)}) / \hat{\varphi}_j \times 100\%$, em que $\hat{\varphi}_{j(I)}$ denota a média a posteriori de φ_j , para $j = 1, \dots, 9$, depois que o conjunto I de observações foi removido. da Tabela 35 pode-se notar que há poucas mudanças relativas na média a posteriori dos parâmetros, mas em geral a significância das estimativas dos parâmetros não muda depois de remover o conjunto I de observações com o nível de significância de 5%. Portanto, não tem-se mudanças inferenciais após remover as observações que nos apresentou os gráficos das medidas de divergência.

Para o modelo Normal-VEG-LL foi estimada a mediana da taxa de sobrevivência para seis pacientes hipotéticos A, B, C, D, E e F com características (covariáveis) dadas na Tabela 33, na qual também são apresentadas as estimativas bayesianas sob perda quadrática e perda absoluta conjuntamente com os intervalos HPD(95%). Da tabela 33 tem-se por exemplo, que para o caso B ou seja os pacientes de sexo feminino e raça hispânica estima-se que tem uma taxa de sobrevivência de 0,901, enquanto que para o caso E isto é os pacientes de sexo masculino e raça não-hispânico negro estima-se que tem uma taxa de sobrevivência da doença de 0,713, esta diferença também pode-se ver na Figura 14, assim como também a distribuição a posteriori da

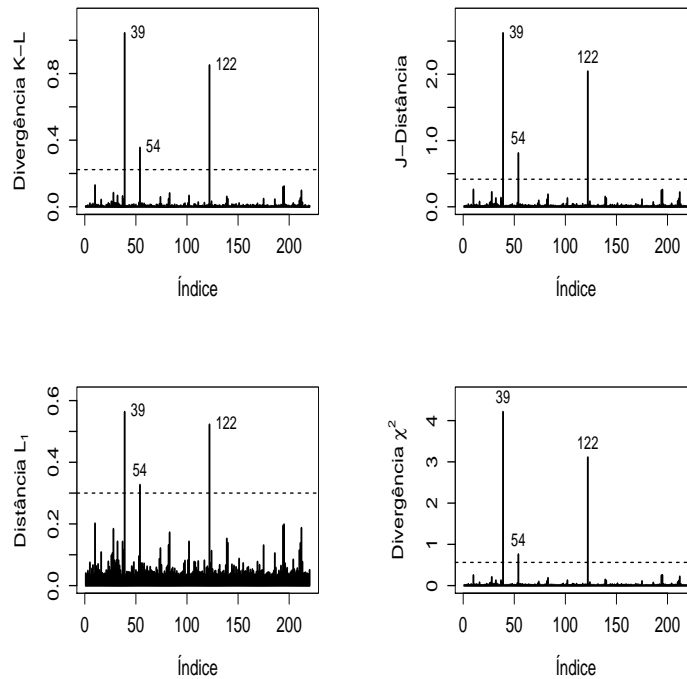


Figura 13 – Gráficos de índices das medidas ψ -divergência para os dados de câncer colorretal

mediana da taxa de sobrevivência dos demais pacientes. Da Tabela 33 e da Figura 14, observa-se que há diferenças significativas na taxa de sobrevivência mediana de pacientes de sexo feminino e masculino, assim como também entre pacientes hispânicos e não-hispânicos.

Tabela 33 – Estimativas da taxa de sobrevivência mediana e o intervalo HPD de 95% de credibilidade para os seis pacientes hipotéticos com câncer colorretal.

Paciente	Sexo	Raça	Taxa de sobrevivência		
			Média	Mediana	Intervalo HPD (95%)
A	Homem	Hispânico	0,857	0,857	(0,847 ; 0,868)
B	Mulher	Hispânico	0,901	0,901	(0,895 ; 0,909)
C	Homem	não-Hispânico branco	0,807	0,806	(0,800 ; 0,812)
D	Mulher	não-Hispânico branco	0,865	0,865	(0,862 ; 0,869)
E	Homem	não-Hispânico negro	0,713	0,713	(0,701 ; 0,726)
F	Mulher	não-Hispânico negro	0,797	0,797	(0,789 ; 0,805)

Tabela 34 – Medidas de ψ -divergência do ajuste do modelo de regressão Normal-VEG-LL para os dados de câncer colorretal

Casos	Taxa de sobrevivência	Sexo	Raça	Estado	$K-L$	J	L_1	χ^2
39	0.890	Homem	Não-Hispânico branco	Dist. Columbia	0,888	1,792	0,489	3,298
54	0.953	Mulher	Hispânico	Georgia	0,299	0,767	0,340	0,731
122	0.71	Mulher	Não-Hispânico negro	Nebraska	0,985	2,000	0,517	3,565

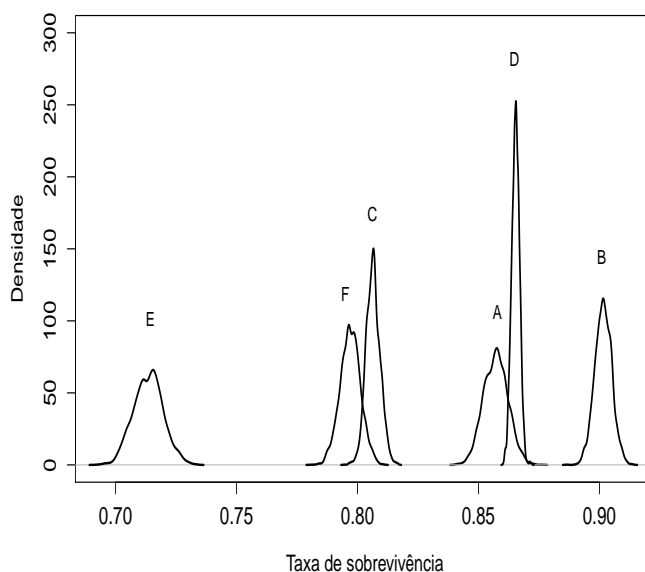


Figura 14 – Gráfico das densidades a posteriori da mediana da taxa de sobrevivência para seis pacientes hipotéticos

Tabela 35 – MRs(%) e seu correspondente intervalo HPD de 95% de credibilidade (LI:limite inferior, LS:limite superior) ajustando o modelo de regressão Normal-VEG aos dados de câncer colorretal sem as observações influentes

sem observação		λ	β_0	β_1	β_{2_1}	β_{2_2}	τ_0	τ_1	τ_{2_1}	τ_{2_2}
	Média	-1,061	1,869	0,396	-0,332	-0,783	3,202	0,703	0,476	-0,490
	LI	-1,617	1,792	0,361	-0,414	-0,870	1,994	0,426	0,135	-1,030
	LS	-0,467	1,952	0,437	-0,260	-0,703	4,228	0,991	0,778	0,026
{39}	MR	25,520	-0,388	-1,237	-3,662	-0,888	15,355	23,425	-35,002	46,923
	LI	-1,377	1,791	0,365	-0,423	-0,871	1,542	0,234	0,303	-0,814
	LS	-0,133	1,953	0,437	-0,264	-0,701	3,947	0,833	0,992	0,339
{54}	MR	7,982	0,462	0,309	2,687	1,775	3,797	0,332	5,282	0,490
	LI	-1,533	1,789	0,358	-0,397	-0,855	2,000	0,422	0,129	-1,003
	LS	-0,364	1,935	0,434	-0,248	-0,688	4,207	0,997	0,716	0,023
{122}	MR	-19,361	0,515	-1,673	0,775	1,899	-11,254	-19,873	15,746	-17,087
	LI	-1,909	1,788	0,361	-0,401	-0,850	2,462	0,518	0,090	-1,096
	LS	-0,704	1,938	0,438	-0,258	-0,691	4,756	1,148	0,740	-0,041
{39,54}	MR	39,249	0,176	-0,237	0,382	1,207	22,413	26,980	-32,498	54,222
	LI	-1,274	1,791	0,360	-0,410	-0,856	1,148	0,194	0,259	-0,842
	LS	0,057	1,936	0,432	-0,258	-0,692	3,698	0,841	0,948	0,321
{39,122}	MR	3,671	0,147	-2,583	-2,303	1,106	2,306	2,327	-15,079	21,673
	LI	-1,661	1,782	0,370	-0,421	-0,862	1,886	0,338	0,178	-0,933
	LS	-0,354	1,940	0,445	-0,263	-0,692	4,379	1,003	0,878	0,223
{54,122}	MR	-14,434	1,110	-1,305	4,638	3,545	-9,486	-19,772	21,658	-19,417
	LI	-1,894	1,781	0,362	-0,387	-0,831	2,254	0,531	0,057	-1,108
	LS	-0,584	1,932	0,439	-0,248	-0,678	4,708	1,186	0,691	-0,022
{39,54,122}	MR	18,133	0,621	-1,660	1,259	2,650	9,767	7,705	-14,224	32,620
	LI	-1,557	1,791	0,365	-0,398	-0,840	1,583	0,310	0,204	-0,910
	LS	-0,203	1,934	0,440	-0,261	-0,684	4,150	0,977	0,872	0,218

4.7 Conclusão

Neste Capítulo, foi construído o modelo de regressão Normal-VEG sob o enfoque bayesiano na qual a estimação dos parâmetros do modelo foi utilizando os métodos de simulação Monte Carlo via Cadeias de Markov (MCMC), especificamente o algoritmo de Metropolis Hasting. Além disso, foram estudadas análises de diagnóstico bayesiano sob varias medidas de divergência para encontrar possíveis pontos influentes na distribuição a posteriori. Também foi conduzido um estudo de simulação para avaliar o comportamento das estimativas bayesianas dos parâmetros do modelo de regressão Normal-VEG. O qual mostrou que as estimativas reproduzem o verdadeiro valor dos parâmetros do modelo à medida que o tamanho da amostra aumenta. Finalmente o modelo foi ajustado aos dados de câncer colorretal, mas desta vez usando a metodologia bayesiana, e obteve-se resultados similares à metodologia clássica usada.

CONSIDERAÇÕES FINAIS E PROPOSTAS FUTURAS

Neste trabalho foi introduzido um novo modelo para a modelagem de dados no intervalo $(0,1)$ como uma alternativa aos modelos já existentes, o qual foi denominado de Modelo Normal-VEG. Para o modelo proposto foram estudadas algumas de suas propriedades, assim como também foram desenvolvidos os procedimentos inferenciais baseadas na teoria de máxima verossimilhança e mediante um estudo de simulação foram verificadas que os EMV dos parâmetros do modelo satisfazem as propriedades assintóticas. Logo o modelo proposto foi estendido para o modelo de regressão, em que as variáveis regressoras estão relacionados com a mediana da variável resposta por meio de uma adequada função de ligação. Para o modelo de regressão foram desenvolvidos procedimentos inferências baseados na teoria de máxima verossimilhança e verificados as propriedades assintóticas dos EMV dos parâmetros do modelo de regressão Normal-VEG homocedástico e heterocedástico mediante um estudo de simulação. O modelo de regressão Normal-VEG também foi desenvolvido no enfoque bayesiano. Foi utilizado o método de simulação Monte Carlo Via Cadeias de Markov para obter a distribuição a posteriori dos parâmetros. Métodos de diagnóstico de influencia foram desenvolvidas neste enfoque para avaliar a robustez do modelo. A utilidade do modelo proposto foi avaliada mediante aplicações a dados reais, onde se mostraram que o modelo proposto é um modelo alternativo para modelagem de dados no intervalo unitário $(0, 1)$.

Como propostas futuras que poderiam ser desenvolvidos:

- Estender o modelo proposto para o caso multivariado;
- Propor uma classe de modelos sob outras transformações diferentes da função quantil da distribuição VEG;
- Desenvolver a metodologia de análise de diagnóstico nas linhas de (COOK, 1986), para avaliar a robustez do modelo desde uma perspectiva clássica.

REFERÊNCIAS

- ANDERSON, T. W. **Statistical inference in elliptically contoured and related distributions**. [S.l.]: Allerton Pr, 1990. Citado na página 24.
- BARNDORFF-NIELSEN, O. E.; JØRGENSEN, B. Some parametric models on the simplex. **Journal of Multivariate Analysis**, Elsevier, v. 39, n. 1, p. 106–116, 1991. Citado na página 23.
- BAYES, C.; BAZAN, J. L.; CASTRO, M. de. A quantile parametric mixed regression model for bounded response variables. **Statistics and its interface**, Elsevier, v. 10, n. 3, p. 483–493, 2017. Citado na página 23.
- BOZDOGAN, H. Model selection and akaike’s information criterion (aic): The general theory and its analytical extensions. **Psychometrika**, Springer, v. 52, n. 3, p. 345–370, 1987. Citado nas páginas 41 e 64.
- BROOKS, S.; SMITH, J.; VEHTARI, A.; PLUMMER, M.; STONE, M.; ROBERT, C. P.; TITTERINGTON, D.; NELDER, J.; ATKINSON, A.; DAWID, A. *et al.* Discussion on the paper by spiegelhalter, best, carlin and van der linde. **Journal of the Royal Statistical Society. Series B: Statistical Methodology**, Wiley-Blackwell, v. 64, n. 4, p. 616–639, 2002. Citado na página 71.
- CANCHO, V. G.; ORTEGA, E. M.; PAULA, G. A. On estimation and influence diagnostics for log-birnbaum–saunders student-t regression models: Full bayesian analysis. **Journal of Statistical Planning and Inference**, Elsevier, v. 140, n. 9, p. 2486–2496, 2010. Citado na página 72.
- CARLIN, B. P.; LOUIS, T. A. **Bayes and empirical Bayes methods for data analysis**. [S.l.]: Chapman and Hall/CRC, 2010. Citado na página 71.
- CARRASCO, J. M.; FERRARI, S. L.; ARELLANO-VALLE, R. B. Errors-in-variables beta regression models. **Journal of Applied Statistics**, Taylor & Francis, v. 41, n. 7, p. 1530–1547, 2014. Citado na página 23.
- COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society, Series B**, v. 48, p. 133–169, 1986. Citado nas páginas 72 e 85.
- COOK, R. D.; WEISBERG, S. **Residuals and Influence in Regression**. Boca Raton, FL: Chapman & Hall/CRC, 1982. Citado na página 72.
- COWLES, M. K.; CARLIN, B. P. Markov chain Monte Carlo convergence diagnostics: A comparative review. **Journal of the American Statistical Association**, v. 91, p. 883–904, 1996. Citado nas páginas 74 e 78.
- DANIEL, W. W. *et al.* **Applied nonparametric statistics**. [S.l.]: Houghton Mifflin, 1978. Citado na página 66.
- DEY, D.; BIRMIWAL, L. R. Robust bayesian analysis using divergence measures. **Statistics & Probability Letters**, Elsevier, v. 20, p. 287–294, 1994. Citado na página 72.

- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. Citado nas páginas 66 e 80.
- FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, Taylor & Francis, v. 31, n. 7, p. 799–815, 2004. Citado na página 23.
- FIGUEROA-ZÚÑIGA, J. I.; ARELLANO-VALLE, R. B.; FERRARI, S. L. Mixed beta regression: A bayesian perspective. **Computational Statistics & Data Analysis**, Elsevier, v. 61, p. 137–147, 2013. Citado na página 23.
- GEISSER, S.; EDDY, W. F. A predictive approach to model selection. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 74, n. 365, p. 153–160, 1979. Citado na página 71.
- GELFAND, A. E.; DEY, D. K.; CHANG, H. **Model determination using predictive distributions with implementation via sampling-based methods**. [S.l.], 1992. Citado na página 71.
- HINKLEY, D. V.; COX, D. **Theoretical statistics**. [S.l.]: Chapman and Hall/CRC, 1979. Citado na página 34.
- JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quarterly Journal of the Royal Meteorological Society**, Wiley Online Library, v. 81, n. 348, p. 158–171, 1955. Citado na página 25.
- JOHNSON, N. L. Systems of frequency curves generated by methods of translation. **Biometrika**, JSTOR, v. 36, n. 1/2, p. 149–176, 1949. Citado na página 24.
- KENNEY, J. F.; KEEPING, E. S. **Mathematics of statistics-part one**. D. Van Nostrand Company Inc.; Toronto; New York; London, 1954. Citado na página 32.
- KRYSICKI, W. On some new properties of the beta distribution. **Statistics & probability letters**, Elsevier, v. 42, n. 2, p. 131–137, 1999. Citado na página 23.
- KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. **Journal of Hydrology**, Elsevier, v. 46, n. 1-2, p. 79–88, 1980. Citado na página 23.
- LEMONTE, A. J.; BAZÁN, J. L. New class of johnson sb distributions and its associated regression model for rates and proportions. **Biometrical Journal**, Wiley Online Library, v. 41, p. 727–746, 2016. Citado nas páginas 23, 24, 41, 44, 65 e 79.
- OSPINA, R.; FERRARI, S. L. Inflated beta distributions. **Statistical Papers**, Springer, v. 51, n. 1, p. 111, 2010. Citado na página 23.
- _____. A general class of zero-or-one inflated beta regression models. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 6, p. 1609–1623, 2012. Citado na página 23.
- PARDO, L. **Statistical Inference Based on Divergence Measures**. Boca Raton, FL: Chapman & Hall/CRC, 2006. Citado na página 72.
- PENG, F.; DEY, D. K. Bayesian analysis of outlier problems using divergence measures. **Canadian Journal of Statistics**, Wiley Online Library, v. 23, n. 2, p. 199–213, 1995. Citado na página 73.

QIU, Z.; SONG, P. X.-K.; TAN, M. Simplex mixed-effects models for longitudinal proportional data. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 35, n. 4, p. 577–596, 2008. Citado na página 23.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2015. Disponível em: <<https://www.R-project.org/>>. Citado nas páginas 31, 63 e 78.

SCHWARZ, G. *et al.* Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. Citado na página 41.

SIEGEL, R.; DESANTIS, C.; JEMAL, A. Colorectal cancer statistics, 2014. **CA: a cancer journal for clinicians**, Wiley Online Library, v. 64, n. 2, p. 104–117, 2014. Citado nas páginas 41 e 63.

SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. **Computational Statistics & Data Analysis**, Elsevier, v. 54, n. 2, p. 348–366, 2010. Citado na página 23.

SMITHSON, M.; SHOU, Y. Cdf-quantile distributions for modelling random variables on the unit interval. **British Journal of Mathematical and Statistical Psychology**, Wiley Online Library, v. 70, n. 3, p. 412–438, 2017. Citado na página 23.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002. Citado na página 71.

TOWNSEND, J.; COLONIUS, H. Variability of the max and min statistic: A theory of the quantile spread as a function of sample size. v. 70, p. 759–772, 02 2005. Citado na página 44.

WATANABE, S. A widely applicable bayesian information criterion. **Journal of Machine Learning Research**, v. 14, n. Mar, p. 867–897, 2013. Citado na página 64.

WEISS, R. An approach to bayesian sensitivity analysis. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 58, n. 4, p. 739–750, 1996. Citado na página 73.

FUNÇÃO ESCORE NORMAL-VEG

Resolvendo a equação $\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \gamma} \Big|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} = 0$, obte-se uma expressão para $\hat{\gamma}$ da seguinte maneira

$$n\hat{\gamma} = -\hat{\delta} \sum_{i=1}^n \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}}$$

na qual tem-se que

$$\hat{\gamma} = -\frac{\hat{\delta}}{n} \sum_{i=1}^n \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}}$$

Tem-se também que $\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \gamma^2} = -n$, que é menor que zero o que garante a unicidade do estimador. Logo a função escore em relação a δ é dada por

$$\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \delta} = \frac{n}{\delta} - \gamma \sum_{i=1}^n \frac{[-\log(y_i)]^{-\lambda} - 1}{\lambda} - \delta \sum_{i=1}^n \left\{ \frac{[-\log(y_i)]^{-\lambda} - 1}{\lambda} \right\}^2.$$

Logo, fazendo $\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \delta} \Big|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} = 0$, tem-se que

$$\begin{aligned} \frac{n}{\hat{\delta}} &= \hat{\gamma} \sum_{i=1}^n \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}} + \hat{\delta} \sum_{i=1}^n \left\{ \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}} \right\}^2 \\ &= -\frac{\hat{\delta}}{n} \sum_{i=1}^n \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}} \sum_{i=1}^n \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}} + \hat{\delta} \sum_{i=1}^n \left\{ \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}} \right\}^2 \\ &= \hat{\delta} \left\{ \sum_{i=1}^n \left\{ \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}} \right\}^2 - \frac{\left\{ \sum_{i=1}^n \frac{[-\log(y_i)]^{-\hat{\lambda}} - 1}{\hat{\lambda}} \right\}^2}{n} \right\}. \end{aligned}$$

Assim, é obtido

$$\hat{\delta} = \sqrt{n \left\{ \sum_{i=1}^n \left\{ \frac{[-\log(y_i)]^{-\hat{\lambda} - 1}}{\hat{\lambda}} \right\}^2 - \frac{\left\{ \sum_{i=1}^n \frac{[-\log(y_i)]^{-\hat{\lambda} - 1}}{\hat{\lambda}} \right\}^2}{n} \right\}^{-1}}$$

Também tem-se $\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \delta^2} = -\frac{n}{\delta^2} - \sum_{i=1}^n \left\{ \frac{[-\log(y_i)]^{-\lambda - 1}}{\lambda} \right\}^2$, que é negativo.

E em relação a λ tem-se que a função escore é dada por

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} &= \frac{\gamma \delta}{\lambda^2} \sum_{i=1}^n \{[-\log(y_i)]^{-\lambda} - 1\} - \frac{\gamma \delta}{\lambda} \sum_{i=1}^n [\log[-\log(y_i)]] [\log(y_i)]^{-\lambda} - \frac{2\delta^2}{\lambda^3} \sum_{i=1}^n \{[-\log(y_i)]^{-\lambda} - 1\}^2 \\ &\quad - \frac{2\delta^2}{\lambda^2} \sum_{i=1}^n \{[-\log(y_i)]^{-\lambda} - 1\} \log[-\log(y_i)] [-\log(y_i)]^{-\lambda} - \log[-\log(y_i)]^{-\lambda}. \end{aligned}$$

FUNÇÃO ESCORE PARA O MODELO DE REGRESSÃO

Seja

$$\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \left(\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)^{\top}, \mathbf{U}_{\boldsymbol{\tau}}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)^{\top}, \mathbf{U}_{\lambda}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) \right)^{\top},$$

os componentes do vetor $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ são obtidas da seguinte forma

$$\frac{\partial \ell(\boldsymbol{\varphi})}{\partial \beta_r} = \sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\varphi})}{\partial v_i} \frac{dv_i}{d\eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_r}, \quad r = 1, \dots, p$$

e

$$\begin{aligned} \frac{\partial \ell_i(\boldsymbol{\varphi})}{\partial v_i} &= \frac{\dot{\phi}\{\delta_i[Q(y_i; \lambda) - Q(v_i; \lambda)]\} [-\delta_i \partial Q(v_i; \lambda)]}{\phi\{\delta_i[Q(y_i; \lambda) - Q(v_i; \lambda)]\} \frac{\partial v_i}{\partial v_i}} \\ &= -\delta_i \{\delta_i[Q(y_i; \lambda) - Q(v_i; \lambda)]\} \frac{-\delta_i}{v_i [-\log(v_i)]^{\lambda+1}} \\ &= \frac{\delta_i \{\delta_i[Q(y_i; \lambda) - Q(v_i; \lambda)]\}}{v_i [-\log(v_i)]^{\lambda+1}} \\ &= \frac{\delta_i w_i}{v_i [-\log(v_i)]^{\lambda+1}}, \end{aligned}$$

em que $w_i = \delta_i[Q(y_i) - Q(v_i)]$. Além disso, $\frac{dv_i}{d\eta_{1i}} = \frac{1}{\frac{d\eta_{1i}}{dv_i}} = \frac{1}{\dot{g}_1(v_i)}$, em que $\dot{g}(x) = \frac{dg(x)}{dx}$ e $\frac{\partial \eta_{1i}}{\partial \beta_r} = x_{ir}$.

Logo, os componentes de $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ tem a forma

$$\frac{\partial \ell(\boldsymbol{\varphi})}{\partial \beta_r} = \sum_{i=1}^n \frac{\delta_i w_i}{v_i [-\log(v_i)]^{\lambda+1}} \frac{x_{ir}}{\dot{g}_1(v_i)}.$$

Matricialmente pode ser expressado por

$$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \mathbf{X}^{\top} \boldsymbol{\Delta} \mathbf{W} \mathbf{G}_1 \mathbf{V},$$

em que \mathbf{X} é a matriz modelo em relação a $\boldsymbol{\beta}$, $\boldsymbol{\Delta} = \text{diag}\{\delta_i\}$, $\mathbf{W} = \text{diag}\{w_i\}$, $\mathbf{G}_1 = \text{diag}\left\{\frac{1}{\dot{g}_1(v_i)}\right\}$, $\mathbf{V} = (V_1, \dots, V_n)^\top$, em que $V_i = \frac{1}{v_i(-\log(v_i))^{\lambda+1}}$, para $i = 1, \dots, n$.

Também pode-se obter os componentes de $\mathbf{U}_\tau(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ da seguinte forma

$$\frac{\partial \ell(\boldsymbol{\varphi})}{\partial \tau_k} = \sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\varphi})}{\partial \delta_i} \frac{d\delta_i}{d\eta_{2i}} \frac{\partial \eta_{2i}}{\partial \tau_k}, \quad k = 1, \dots, q$$

e

$$\begin{aligned} \frac{\partial \ell_i(\boldsymbol{\varphi})}{\partial \delta_i} &= \frac{1}{\delta_i} + \frac{\phi\{\delta_i[Q(y_i; \lambda) - Q(v_i; \lambda)]\}}{\phi\{\delta_i[Q(y_i; \lambda) - Q(v_i; \lambda)]\}} \\ &= \frac{1}{\delta_i} - \delta_i[Q(y_i; \lambda) - Q(v_i; \lambda)][Q(y_i; \lambda) - Q(v_i; \lambda)] \\ &= \frac{1}{\delta_i} - \frac{\delta_i^3[Q(y_i; \lambda) - Q(v_i; \lambda)]^2}{\delta_i} \\ &= \frac{1}{\delta_i} - \frac{w_i^2}{\delta_i}, \end{aligned}$$

em que $\frac{d\delta_i}{d\eta_{2i}} = \frac{1}{\frac{d\eta_{2i}}{d\delta_i}} = \frac{1}{\dot{g}_2(\delta_i)}$, e $\frac{\partial \eta_{2i}}{\partial \tau_k} = s_{ik}$. Logo temos que a expressão para cada componente do $\mathbf{U}_\tau(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda)$ é dado por

$$\frac{\partial \ell(\boldsymbol{\varphi})}{\partial \tau_k} = \sum_{i=1}^n \left(\frac{1}{\delta_i} - \frac{w_i^2}{\delta_i} \right) \frac{s_{ik}}{\dot{g}_2(\delta_i)}, \quad k = 1, \dots, q.$$

Matricialmente pode-se expressar como

$$\mathbf{U}_\tau(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) = \mathbf{S}^\top \mathbf{G}_2 \mathbf{T},$$

em que \mathbf{S} matriz modelo em relação a $\boldsymbol{\tau}$, $\mathbf{G}_2 = \text{diag}\left\{\frac{1}{\dot{g}_2(\delta_i)}\right\}$, $\mathbf{T} = (T_1, \dots, T_n)^\top$ e $T_i = \frac{1}{\delta_i} - \frac{w_i^2}{\delta_i}$. Por último a função escore para λ é obtida derivando $\ell(\boldsymbol{\varphi})$ em relação a λ , e será dada pela seguinte expressão

$$\begin{aligned} U_\lambda(\boldsymbol{\beta}, \boldsymbol{\tau}, \lambda) &= \sum_{i=1}^n \frac{w_i \delta_i}{\lambda} \left\{ [-\log(y_i)]^{-\lambda} \left\{ \frac{1}{\lambda} + \log[-\log(y_i)] \right\} - [-\log(v_i)]^{-\lambda} \left\{ \frac{1}{\lambda} + \log[-\log(v_i)] \right\} \right\} \\ &\quad - \sum_{i=1}^n \log[-\log(y_i)]. \end{aligned}$$

MATRIZ DE INFORMAÇÃO OBSERVADA

O cálculo dos elementos da matriz $J_{\beta\beta}$ é da seguinte maneira

$$\frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^n \frac{\partial}{\partial v_i} \left(\frac{\partial \ell_i(\boldsymbol{\varphi})}{\partial v_i} \frac{dv_i}{d\eta_{1i}} \right) \frac{dv_i}{d\eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_s} x_{ir}$$

ou seja

$$\frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^n \frac{\partial}{\partial v_i} \left\{ \frac{\delta_i w_i}{v_i [-\log(v_i)]^{\lambda+1}} \frac{1}{\dot{g}_1(v_i)} \right\} \frac{\partial v_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_s} x_{ir},$$

em que

$$\begin{aligned} L_1 &= \frac{\partial}{\partial v_i} \left(\frac{\delta_i w_i}{v_i [-\log(v_i)]^{\lambda+1}} \frac{1}{\dot{g}_1(v_i)} \right) \\ &= \frac{\partial}{\partial v_i} \left\{ \frac{\delta_i (\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)])}{v_i [-\log(v_i)]^{\lambda+1}} \right\} \frac{1}{\dot{g}_1(v_i)} + \frac{\delta_i (\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)])}{v_i [-\log(v_i)]^{\lambda+1}} \frac{\partial}{\partial v_i} \frac{1}{\dot{g}_1(v_i)} \\ &= \frac{-\delta_i^2 - \{[-\log(v_i)]^{\lambda+1} - (\lambda+1)[- \log(v_i)]^\lambda\} \delta_i^2 [Q(y_i; \lambda) - Q(v_i; \lambda)]}{\{v_i [-\log(v_i)]^{\lambda+1}\}^2 \dot{g}_1(v_i)} \\ &\quad + \frac{\delta_i (\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)])}{v_i [-\log(v_i)]^{\lambda+1}} \frac{-\dot{g}_1(v_i)}{(\dot{g}_1(v_i))^2} \\ &= \frac{-\delta_i^2 - \{[-\log(v_i)]^{\lambda+1} - (\lambda+1)[- \log(v_i)]^\lambda\} [\delta_i w_i]}{(v_i [-\log(v_i)]^{\lambda+1})^2 \dot{g}_1(v_i)} - \frac{\delta_i w_i \ddot{g}_1(v_i)}{v_i [-\log(v_i)]^{\lambda+1} (\dot{g}_1(v_i))^2} \end{aligned}$$

$$\text{e } \ddot{g}(x) = \frac{d\dot{g}(x)}{dx}.$$

Finalmente, tem-se que

$$\frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^n \left(\frac{-\delta_i^2 - [-\log(v_i)]^\lambda \{[-\log(v_i)] - \lambda - 1\} \delta_i w_i}{(v_i [-\log(v_i)]^{\lambda+1})^2 \dot{g}_1(v_i)} - \frac{\delta_i w_i \ddot{g}_1(v_i)}{v_i [-\log(v_i)]^{\lambda+1} (\dot{g}_1(v_i))^2} \right) \frac{1}{\dot{g}_1(v_i)} x_{ir} x_{is}.$$

Matricialmente é expressado da forma

$$J_{\beta\beta} = \mathbf{X}^\top \mathbf{W}_1 \mathbf{G}_1 \mathbf{X},$$

em que $\mathbf{G} = \text{diag}\{\frac{1}{\dot{g}_1(v_i)}\}$, $\mathbf{W}_1 = \text{diag}\{W_{1i}\}$, e

$$W_{1i} = \frac{-\delta_i^2 - [-\log(v_i)]^\lambda [-\log(v_i) - \lambda - 1] \delta_i w_i}{(v_i [-\log(v_i)]^{\lambda+1})^2 \dot{g}_1(v_i)} - \frac{\delta_i w_i \dot{g}_1(v_i)}{v_i [-\log(v_i)]^{\lambda+1} (\dot{g}_1(v_i))^2}.$$

Também, pode-se obter $\mathbf{J}_{\tau\tau}$ da seguinte maneira

$$\frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \tau_k \partial \tau_l} = \sum_{i=1}^n \frac{\partial}{\partial \delta_i} \left(\frac{\partial \ell_i(\boldsymbol{\varphi})}{\partial \delta_i} \frac{d\delta_i}{d\eta_{2i}} \right) \frac{d\delta_i}{d\eta_{2i}} \frac{\partial \eta_{2i}}{\partial \tau_l} S_{ik}$$

ou seja

$$\frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \tau_k \partial \tau_l} = \sum_{i=1}^n \frac{\partial}{\partial \delta_i} \left(\left[\frac{1}{\delta_i} - \frac{(\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)])^2}{\delta_i} \right] \frac{1}{\dot{g}_2(\delta_i)} \right) \frac{d\delta_i}{d\eta_{2i}} \frac{\partial \eta_{2i}}{\partial \tau_l} S_{ik},$$

em que

$$\begin{aligned} L_2 &= \frac{\partial}{\partial \delta_i} \left(\left[\frac{1}{\delta_i} - \frac{(\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)])^2}{\delta_i} \right] \frac{1}{\dot{g}_2(\delta_i)} \right) \\ &= \frac{\partial}{\partial \delta_i} \left(\frac{1}{\delta_i} - \delta_i [Q(y_i) - Q(v_i; \lambda)]^2 \right) \frac{1}{\dot{g}_2(\delta_i)} + \frac{\partial}{\partial \delta_i} \left(\frac{1}{\dot{g}_2(\delta_i)} \right) \left(\frac{1}{\delta_i} - \delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)]^2 \right) \\ &= \left(-\frac{1}{\delta_i^2} - (Q(y_i; \lambda) - Q(v_i; \lambda))^2 \right) \frac{1}{\dot{g}_2(\delta_i)} + \frac{1 - \delta_i^2 [Q(y_i; \lambda) - Q(v_i; \lambda)]^2 (-\ddot{g}_2(\delta_i))}{\delta_i (\dot{g}_2(\delta_i))^2} \\ &= \frac{-(1 + \delta_i^2 [Q(y_i; \lambda) - Q(v_i; \lambda)]^2) \dot{g}_2(\delta_i) - \delta_i (1 - \delta_i^2 [Q(y_i; \lambda) - Q(v_i; \lambda)]^2) \ddot{g}_2(\delta_i)}{\delta_i^2 (\dot{g}_2(\delta_i))^2} \\ &= \frac{-(1 + w_i^2)}{\delta_i^2 \dot{g}_2(\delta_i)} - \frac{\delta_i (1 - w_i^2) \ddot{g}_2(\delta_i)}{\delta_i^2 (\dot{g}_2(\delta_i))^2}. \end{aligned}$$

Finalmente, tem-se que

$$\frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \tau_k \partial \tau_l} = \sum_{i=1}^n \left(\frac{-(1 + w_i^2)}{\delta_i^2 \dot{g}_2(\delta_i)} - \frac{\delta_i (1 - w_i^2) \ddot{g}_2(\delta_i)}{\delta_i^2 (\dot{g}_2(\delta_i))^2} \right) \frac{1}{\dot{g}_2(\delta_i)} S_{il} S_{ik}.$$

Matricialmente pode-se escrever como

$$\mathbf{J}_{\tau\tau} = \mathbf{S}^\top \mathbf{W}_2 \mathbf{G}_2 \mathbf{S},$$

em que $\mathbf{W}_2 = \text{diag}\{W_{2i}\}$, sendo $W_{2i} = \frac{-(1 + w_i^2)}{\delta_i^2 \dot{g}_2(\delta_i)} - \frac{\delta_i (1 - w_i^2) \ddot{g}_2(\delta_i)}{\delta_i^2 (\dot{g}_2(\delta_i))^2}$.

Além disso, também pode-se calcular a expressão para $\mathbf{J}_{\beta\tau}$, com elementos que são calculados da forma

$$\frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \beta_r \partial \tau_k} = \sum_{i=1}^n \frac{\partial}{\partial \delta_i} \left(\frac{\partial \ell_i(\boldsymbol{\varphi})}{\partial v_i} \right) \frac{d\delta_i}{d\eta_{2i}} \frac{\partial \eta_{2i}}{\partial \tau_k} \frac{dv_i}{d\eta_{1i}} x_{ir},$$

em que

$$\frac{\partial}{\partial \delta_i} \left(\frac{\partial \ell_i(\boldsymbol{\varphi})}{\partial v_i} \right) = \frac{\partial}{\partial \delta_i} \left(\frac{\delta_i \delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)]}{v_i [-\log(v_i)]^{\lambda+1}} \right) = \frac{2\delta_i [Q(y_i; \lambda) - Q(v_i; \lambda)]}{v_i [-\log(v_i)]^{\lambda+1}} = \frac{2w_i}{v_i [-\log(v_i)]^{\lambda+1}}.$$

Assim, cada elemento da matriz $\mathbf{J}_{\beta\tau}$ terá a seguinte expressão

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \beta_r \partial \tau_k} &= \sum_{i=1}^n \left\{ \frac{2w_i}{v_i [-\log(v_i)]^{\lambda+1}} \right\} \frac{1}{\dot{g}_2(\delta_i)} S_{ik} \frac{1}{\dot{g}_1(v_i)} x_{ir} \\ &= \sum_{i=1}^n \left\{ \frac{2w_i}{v_i [-\log(v_i)]^{\lambda+1}} \right\} \frac{1}{\dot{g}_2(\delta_i)} \frac{1}{\dot{g}_1(v_i)} x_{ir} S_{ik}. \end{aligned}$$

Matricialmente pode-se expressar como

$$\mathbf{J}_{\beta\tau} = \mathbf{X}^T \mathbf{W}_3 \mathbf{G}_1 \mathbf{G}_2 \mathbf{S},$$

em que $\mathbf{W}_3 = \text{diag}\{W_{3i}\}$, sendo $W_{3i} = \frac{2w_i}{v_i[-\log(v_i)]^{\lambda+1}}$.

Também, tem-se que os elementos para $\mathbf{J}_{\beta\lambda}$ é dado por

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \beta_r \partial \lambda} &= \sum_{i=1}^n \frac{\delta_i}{\lambda} \left\{ \frac{[-\log(v_i)]^{-\lambda} \left(\frac{1}{\lambda} + \log[-\log(v_i)] \right) - [-\log(y_i)]^{-\lambda} \left(\frac{1}{\lambda} + \log[-\log(y_i)] \right)}{v_i[-\log(v_i)]^{\lambda+1}} \right\} \frac{x_{ir}}{\dot{g}_1(v_i)} \\ &\quad - \sum_{i=1}^n \frac{w_i(\lambda+1)[- \log(v_i)]^\lambda}{\{v_i[-\log(v_i)]^{\lambda+1}\}^2} \frac{x_{ir}}{\dot{g}_1(v_i)}. \end{aligned}$$

Os elementos de $\mathbf{J}_{\tau\lambda}$ é dado por

$$\frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \tau_k \partial \lambda} = \sum_{i=1}^n \frac{2w_i}{\lambda} \left\{ [-\log(y_i)]^{-\lambda} \left\{ \frac{1}{\lambda} + \log[-\log(y_i)] \right\} - [-\log(v_i)]^{-\lambda} \left\{ \frac{1}{\lambda} + \log[-\log(v_i)] \right\} \right\} \frac{s_{ik}}{\dot{g}_2(\delta_i)}$$

$$\frac{\partial^2 \ell(\boldsymbol{\varphi})}{\partial \lambda^2} = \sum_{i=1}^n -\delta_i^2 \left\{ [\dot{Q}(y_i; \lambda) - \dot{Q}(v_i; \lambda)]^2 + [Q(y_i; \lambda) - Q(v_i; \lambda)][\ddot{Q}(y_i; \lambda) - \ddot{Q}(v_i; \lambda)] \right\},$$

em que $\dot{Q}(\cdot) = dQ(\cdot)/d\lambda$, e $\ddot{Q}(\cdot) = d^2Q(\cdot)/d\lambda^2$. Assim, temos que

$$\dot{Q}(y_i; \lambda) - \dot{Q}(v_i; \lambda) = \frac{[-\log(y_i)]^{-\lambda}}{\lambda} \left\{ \frac{-1}{\lambda} - \log[-\log(y_i)] \right\} - \frac{[-\log(v_i)]^{-\lambda}}{\lambda} \left\{ \frac{-1}{\lambda} - \log[-\log(v_i)] \right\}$$

e

$$\begin{aligned} \ddot{Q}(y_i; \lambda) - \ddot{Q}(v_i; \lambda) &= \frac{[-\log(y_i)]^{-\lambda}}{\lambda} \left\{ \frac{2}{\lambda^2} + \frac{2}{\lambda} \log[-\log(y_i)] + \{\log[-\log(y_i)]\}^2 \right\} \\ &\quad - \frac{[-\log(v_i)]^{-\lambda}}{\lambda} \left\{ \frac{2}{\lambda^2} + \frac{2}{\lambda} \log[-\log(v_i)] + \{\log[-\log(v_i)]\}^2 \right\}. \end{aligned}$$