



UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
UNIDADE ACADÊMICA (UA) FACULDADE DE
ADMINISTRAÇÃO, CIÊNCIAS CONTÁBEIS
E CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS GRADUAÇÃO EM ECONOMIA
APLICADA

IHURI NUNES BARROSO

Inferência estatística via *bootstrap* no modelo de regressão Gama Unitária

Programa de Pós-Graduação em Economia
Orientadora: Prof^a. Dr^a. Tatiane Ferreira do Nascimento Melo da Silva

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da Capes

Goiânia 2022



UNIVERSIDADE FEDERAL DE GOIÁS
FACULDADE DE ADMINISTRAÇÃO, CIÊNCIAS CONTÁBEIS E CIÊNCIAS ECONÔMICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Ihuri Nunes Barroso

3. Título do trabalho

Inferência estatística via *bootstrap* no modelo de regressão Gama Unitária

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Tatiane Ferreira Do Nascimento Melo Da Silva, Professor do Magistério Superior**, em 16/09/2022, às 18:38, conforme horário oficial de Brasília, com



fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **IHURI NUNES BARROSO, Discente**, em 21/09/2022, às 15:59, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3196790** e o código CRC **62E2C165**.

IHURI NUNES BARROSO

Inferência estatística via *bootstrap* no modelo de regressão Gama Unitária

Dissertação apresentada ao Programa de Pós-Graduação em Economia Aplicada, da Faculdade de Administração, Ciências Contábeis e Ciências Econômicas, da Universidade Federal de Goiás (UFG), como requisito para a obtenção do título de Mestre em Economia.
Área de concentração: Economia Aplicada.
Linha de pesquisa: Organizações e Mercado.
Orientadora: Prof^a. Dr^a. Tatiane Ferreira do Nascimento Melo da Silva

Goiânia 2022

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Barroso, Ihuri Nunes

Inferência estatística via bootstrap no modelo de regressão Gama Unitária [manuscrito] / Ihuri Nunes Barroso. - 2022.
xlvii, 46 f.

Orientador: Profa. Dra. Tatiane Ferreira do Nascimento Melo da Silva.

Dissertação (Mestrado) - Universidade Federal de Goiás, Faculdade de Administração, Ciências Contábeis e Ciências Econômicas (FACE), Programa de Pós-Graduação em Economia, Goiânia, 2022.

Bibliografia. Anexos. Apêndice.

Inclui siglas, abreviaturas, tabelas, algoritmos, lista de figuras, lista de tabelas.

1. Bootstrap. 2. Energia fotovoltaica. 3. Inferência Estatística. 4. Regressão Gama Unitária. 5. Variáveis socioeconômicas. I. Silva, Tatiane Ferreira do Nascimento Melo da, orient. II. Título.



UNIVERSIDADE FEDERAL DE GOIÁS

FACULDADE DE ADMINISTRAÇÃO, CIÊNCIAS CONTÁBEIS E CIÊNCIAS ECONÔMICAS

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 04 da sessão de Defesa de Dissertação de Ihuri Nunes Barroso, que confere o título de Mestre em Economia, na área de concentração em Economia Aplicada.

Aos dezesseis dias do mês de setembro do ano de dois mil e vinte e dois, a partir das 10h00min, por videoconferência, realizou-se a sessão pública de Defesa de Dissertação intitulada “ Inferência estatística via *bootstrap* no modelo de regressão Gama Unitária”. Os trabalhos foram instalados pela Orientadora, Professora Doutora Tatiane Ferreira do Nascimento Melo da Silva (PPGECON/UFG) com a participação dos demais membros da Banca Examinadora: Professor Doutor Eder Angelo Milani (IME/UFG) membro titular externo; Professor Doutor Sandro Eduardo Monsueto (PPGECON/UFG) membro titular interno. Durante a arguição os membros da banca não fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato aprovado pelos seus membros. Proclamados os resultados pela Professora Doutora Tatiane Ferreira do Nascimento Melo da Silva, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos dezesseis dias do mês de setembro do ano de dois mil e vinte e dois.

TÍTULO SUGERIDO PELA BANCA

Documento assinado eletronicamente por **Tatiane Ferreira Do Nascimento Melo Da Silva, Professor do Magistério Superior**, em 16/09/2022, às 13:49, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eder Angelo Milani, Professor do Magistério Superior**, em 16/09/2022, às 14:37, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sandro Eduardo Monsueto, Professor do Magistério Superior**, em 16/09/2022, às 14:53, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3165656** e o código CRC **121CC003**.

Este trabalho é dedicado aos meus pais, a Deus, à minha orientadora e a todos os demais que fizeram parte da minha vida e me incentivaram nesta trajetória.

Agradecimentos

Agradeço primeiro a Deus por ter me mantido na trilha certa durante este projeto de pesquisa com saúde e forças para chegar até o final.

Sou grato aos meus pais José Carmênio Barroso Júnior e Marinalva Nunes Barroso pelo apoio que sempre me deram durante toda a minha vida, inclusive nos estudos, me incentivando a seguir a trilha da pesquisa.

Deixo um agradecimento especial à minha orientadora Tatiane Ferreira do Nascimento Melo da Silva pelo incentivo, pela paciência e dedicação do seu tempo ao meu projeto de pesquisa.

Também quero agradecer à Universidade Federal de Goiás, a todos os professores do meu curso pela elevada qualidade do ensino oferecido e à CAPES pelo empenho e incentivo à pesquisa.

Resumo

Em Economia, existem muitas situações envolvendo dados restritos ao intervalo $(0,1)$, ou seja, dados de taxas e proporções, e há modelos que se adequam melhor à esta situação, como, por exemplo, o modelo de regressão Gama Unitária. Porém, quando o tamanho da amostra é pequeno, ou mesmo moderado, a Inferência Estatística destes modelos fica comprometida. Os estimadores, em geral, tendem a se tornar mais viesados e as estatísticas de testes conduzem a testes menos precisos. Assim, é necessário usar ferramentas que são capazes de corrigir o viés dos estimadores e as estatísticas de testes, como o método de *bootstrap*. Neste trabalho, propomos simulações de Monte Carlo, via *bootstrap*, que resolvem os problemas supracitados. Além disso, estudamos variáveis socioeconômicas que impactam a geração de energia através de sistemas fotovoltaicos, usando o modelo de regressão Gama Unitária e Inferência Estatística via *bootstrap*.

Palavras-chave: *Bootstrap*; Energia fotovoltaica, Inferência Estatística; Regressão Gama-Unitária, Variáveis socioeconômicas.

Abstract

In Economics, there are many situations involving data restricted to the range $(0,1)$, that is, data of rates and proportions, and there are models that are better suited to this situation, such as the Unit Gamma regression model . However, when the sample size is small, or even moderate, the Statistical Inference of these models is compromised. Estimators, in general, tend to become more biased and test statistics lead to less accurate tests. Thus, it is necessary to use tools that are able to correct the bias of estimators and test statistics, such as the method of bootstrap. In this work, we propose Monte Carlo simulations, via bootstrap, which solve the aforementioned problems. In addition, we study socioeconomic variables that impact energy generation through photovoltaic systems, using the Unit Gamma regression model and Statistical Inference via it bootstrap.

Keywords: *bootstrap*; Gamma-Unit Regression; Statistical inference.

Lista de Tabelas

3.1	Estimativas do parâmetros: $q = 2$, $p = 3, 4$ e diferentes valores de n ; $\beta_1 = 1$, $\beta_2 = \beta_3 = 0$ e $\phi = 10$	14
3.2	Estimativas do parâmetros: $q = 3$, $p = 4$ e diferentes valores de n , $\beta_1 = 1$, $\beta_2 = \beta_3 = \beta_4 = 0$ e $\phi = 10$	15
3.3	Taxas de rejeição nula: $q = 2$, $p = 3, 4$ e diferentes valores de n	16
3.4	Taxas de rejeição nula: $q = 3$, $p = 4$ e diferentes valores de n	17
4.1	Dados usados na aplicação.	22

Sumário

Lista de Tabelas	v
1 Introdução	1
2 O modelo de regressão Gama Unitária	4
3 Inferência estatística via <i>bootstrap</i>	8
3.1 Correção de viés	8
3.2 Testes de hipótese	10
3.3 Resultados numéricos	12
4 Aplicação	18
4.1 Descrição dos dados	18
4.2 Modelagem	21
5 Conclusões	24
A Vetor escore e matriz de informação de Fisher	26
A.1 Densidade da Gama Unitária	26
A.2 Vetor escore	28
A.3 Matriz de informação de Fisher	34
Referências Bibliográficas	42

Capítulo 1

Introdução

Em Economia existem situações práticas em que o interesse consiste em modelar dados restritos no intervalo $(0,1)$, como por exemplo, taxa de crescimento econômico dos países, proporção de pessoas que se encontram fora da linha de pobreza de um determinado país ou até mesmo o percentual de potência instalada da geração distribuída em um determinado local. Nestes casos, é natural usar modelos para taxas e proporções. Um modelo muito utilizado na literatura para dados no intervalo unitário padrão é o modelo de regressão Beta, proposto por Ferrari e Cribari-Neto (2004). Neste modelo, os autores incorporam variáveis explicativas na média da distribuição da variável resposta através de uma função de ligação. Um modelo alternativo, para dados no intervalo $(0,1)$, é o modelo de regressão Gama Unitária. Mousa et al. (2016) verificaram que várias propriedades da distribuição Gama Unitária são semelhantes às da distribuição Beta. Rocha (2020) apresenta gráficos onde é possível verificar que quando a média da variável resposta fica perto de 1 (um) há fortes indícios de que o modelo de regressão Gama Unitária apresenta uma modelagem melhor do que quando comparado ao modelo Beta, já que para o primeiro caso, nesta situação, temos variância menor que o do último modelo. Na literatura há vários trabalhos acerca de modelos para taxas e proporções, principalmente para o modelo Beta, a saber: Espinheira et al. (2008) realizaram um estudo de diagnóstico; Simas et al. (2010) propuseram uma estrutura não linear para a regressão e apresentaram correções de vieses para os estimadores dos parâmetros, quando o tamanho da amostra é pequeno; Carrasco et al. (2014) propuseram um modelo Beta com erros nas variáveis. Outros trabalhos para dados

envolvendo taxas e proporções são: Mazucheli et al. (2019) que propuseram, transformando a distribuição Lindley (LINDLEY, 1958), a distribuição Lindley-Unitária. Guedes et al. (2020) e Rocha et al. (2021) propuseram refinamentos para estatística da razão de verossimilhanças e um estudo de análise de resíduos e influência local para o modelo de regressão Gama Unitária, respectivamente.

Em modelos estatísticos, os estimadores de máxima verossimilhança (EMVs), geralmente, são viesados quando a amostra é pequena ou mesmo de tamanho moderado. Em geral, o viés deste estimador é de ordem n^{-1} , com n sendo o tamanho da amostra. Para contornar este problema podemos usar uma correção no viés destes estimadores. Um método analítico muito conhecido na literatura para correção de viés é o de Cox e Snell (1968). Mas, em alguns modelos, a obtenção das quantidades usadas neste procedimento é difícil ou até impossível. Neste caso, uma alternativa é usar o método de reamostragem *bootstrap* (EFRON, 1979). Além dos EMVs serem viesados, quando a amostra é pequena, temos que as aproximações assintóticas de primeira ordem dos testes estatísticos clássicos podem ser pouco precisas. Em geral, esse é o caso do teste da razão de verossimilhanças (NEYMAN e PERSON, 1928), teste de Wald (WALD, 1943), teste Escore (RAO, 1948) e teste Gradiente (TERRELL, 2002). As estatísticas dos testes supracitados podem ser transformadas para que, sob a hipótese nula, tenham distribuições mais próximas das distribuição qui-quadrado do que as estatísticas não modificadas. Na literatura existem vários métodos analíticos para refinamentos nas estatísticas de testes, como, por exemplo, Bartlett (1937), Barndorff-Nielsen (1991) e Skovgaard (2001). Mas, nestes procedimentos também precisamos de quantidades difíceis de serem obtidas analiticamente e então, uma alternativa é, também, usar método *bootstrap* para encontrar as estatísticas modificadas.

A importância deste trabalho se dá pelo fato de se tratar da correção de um modelo importante para estudar variáveis no intervalo entre $(0,1)$, mesmo quando se está lidando com amostras de tamanho pequeno, como acontece em várias pesquisas importantíssimas. Por isso, o objetivo deste trabalho seguiu duas vertentes importantes na Inferência Estatística para o modelo de regressão Gama Unitária, quando o tamanho da amostra é pequeno. A primeira foi a implementação, no software R (R Core Team, 2021), da correção de viés via *bootstrap* dos EMVs, de modo a reduzir os vieses destes estimadores. Em um segundo momento, apresentamos

versões modificadas via *bootstrap* das estatísticas da razão de verossimilhanças, Wald, Escore e Gradiente para testes de hipóteses no modelo supracitado. Não há na literatura estudos de simulações como o que está proposto neste trabalho. Além dos resultados numéricos, apresentamos uma aplicação a dados reais, em que usamos as estimativas e os testes corrigidos, já que estes são mais precisos quando temos um tamanho amostral pequeno. A variável dependente, neste caso, é o percentual da potência instalada de geração de energia solar e as independentes são: tarifa da energia de baixa tensão, Índice Firjan de Desenvolvimento Municipal (IFDM), *payback* e taxa de ocupação de algumas capitais do Brasil.

Este trabalho foi estruturado da seguinte forma: no Capítulo 2, apresentamos o modelo de regressão Gama Unitária; no Capítulo 3, apresentamos a inferência estatística baseada no método de *bootstrap*, considerando a correção de viés e os testes de hipóteses. Adicionalmente, neste capítulo, desenvolvemos estudos de simulação relacionados a correção de viés e os testes de hipóteses via *bootstrap*; no Capítulo 4, foi apresentada uma aplicação a dados reais; finalmente, no Capítulo 5 apresentamos as conclusões referentes à pesquisa. Além disso, no Apêndice A, temos o vetor escore, a matriz de informação observada e a matriz esperada de informação de Fisher do modelo de regressão Gama Unitária, já evidenciados no Capítulo 2.

Capítulo 2

O modelo de regressão Gama Unitária

A distribuição Gama foi proposta por Grassia (1977), cuja função de densidade de probabilidade (f.d.p.) é da forma:

$$g(w; \alpha, \phi) = \frac{\alpha^\phi}{\Gamma(\phi)} e^{-\alpha \cdot w} w^{\phi-1}, \quad w > 0, \quad (2.1)$$

com parâmetros $\alpha > 0$ e $\phi > 0$. Para esta distribuição, usamos a notação $W \sim Gama(\alpha, \phi)$. A esperança e variância de W são dados, respectivamente, por: $E(W) = \phi/\alpha$ e $Var(W) = \phi/\alpha^2$. A partir da distribuição Gama podemos obter a distribuição Gama Unitária usando a transformação $w = \log(1/y)$, que implica em $y = e^{-w}$, $0 < y < 1$. No Apêndice A mostramos que, a f.d.p. da Gama Unitária ($Y \sim UG(\alpha, \phi)$), com parâmetros α e ϕ , é dada por:

$$f(y; \alpha, \phi) = \frac{\alpha^\phi}{\Gamma(\phi)} y^{\alpha-1} (-\log y)^{\phi-1}, \quad 0 < y < 1. \quad (2.2)$$

Rocha (2020) apresenta a média, variância e momentos de Y , que são dados, respectivamente, por:

$$E(Y) = \left(\frac{\alpha}{\alpha+1}\right)^\phi, \quad Var(Y) = \left(\frac{\alpha}{\alpha+2}\right)^\phi - \left(\frac{\alpha}{\alpha+1}\right)^{2\phi} \quad \text{e} \quad E(Y^n) = \left(\frac{\alpha}{n+\alpha}\right)^\phi.$$

Detalhes do cálculo de $E(Y)$ é apresentado no Apêndice A.

Aqui, usamos uma reparametrização da f.d.p. em termos da média com

$$\alpha = \frac{\mu^{1/\phi}}{1 - \mu^{1/\phi}}.$$

Então, (2.2) pode ser escrita como

$$f(y; \mu, \phi) = \frac{1}{\Gamma(\phi)} \left[\frac{\mu^{1/\phi}}{1 - \mu^{1/\phi}} \right]^\phi y^{[\mu^{1/\phi}/(1-\mu^{1/\phi})]^{-1}} (-\log y)^{\phi-1}, \quad (2.3)$$

com $0 < y < 1$, $0 < \mu < 1$ e $\phi > 0$. O valor esperado e a variância de Y são, respectivamente, $E(Y) = \mu$ e $Var(Y) = \mu \{[\mu/(2 - \mu^{1/\phi})^\phi] - \mu\}$, veja Rocha (2020).

Considere Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes, em que cada Y_i , com $i = 1, \dots, n$, com média μ_i e precisão ϕ desconhecidas, tem f.d.p dada pela equação (2.3). O modelo de regressão Gama Unitária proposto por Mousa et al. (2016) é dado por

$$g(\mu_i) = \mathbf{X}_i^\top \boldsymbol{\beta} = \sum_{j=1}^p x_{ij} \beta_j = \beta_1 + \beta_2 x_{i2} + \dots + \beta_p x_{ip} = \eta_i, \quad (2.4)$$

em que $g : (0,1) \rightarrow \Re$ é uma função de ligação estritamente monótona e duas vezes diferenciável, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ é o vetor p -dimensional de parâmetros desconhecidos ($p < n$) e $\mathbf{X}_i^\top = (1, x_{i2}, \dots, x_{ip})$ são observações de p covariáveis conhecidas.

O logaritmo da função de verossimilhança pode ser escrito na forma

$$\ell(\boldsymbol{\beta}, \phi) = \sum_{i=1}^n \ell_i(\mu_i, \phi), \quad (2.5)$$

onde

$$\ell_i(\mu_i, \phi) = \phi \log(\alpha_i) - \log[\Gamma(\phi)] + (\alpha_i - 1) \log(y_i) + (\phi - 1) \log[-\log(y_i)],$$

com $\mu_i = g^{-1}(\eta_i) = h(\eta_i)$.

O vetor escore $(p + 1)$ -dimensional, obtido derivando a função log-verossimilhança (2.5)

com relação a $\boldsymbol{\beta}$ e ϕ , é dado por

$$U(\boldsymbol{\beta}, \phi) = \begin{bmatrix} U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi) \\ U_{\phi}(\boldsymbol{\beta}, \phi) \end{bmatrix},$$

onde $U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi) = X^{\top} T \mathbf{s}$, com X sendo a matriz ($n \times p$) de covariadas, $T = \text{diag}\{h'(\eta_1), \dots, h'(\eta_n)\}$ e o i -ésimo elemento do vetor \mathbf{s} é dado por

$$s_i = \frac{\alpha_i}{\mu_i^{1/\phi+1}} + \frac{\alpha_i^2 \log(y_i)}{\phi \mu_i^{1/\phi+1}}.$$

Denotamos $h'(\eta_i)$ como a derivada de primeira ordem de $h(\eta_i) = g^{-1}(\eta_i)$, com $i = 1, 2, \dots, n$. Se a função de ligação é logito, ou seja, $g(\mu_i) = \log(\mu_i/(1 - \mu_i))$, então

$$h(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}}.$$

Logo,

$$h'(\eta_i) = \frac{e^{\eta_i}}{(1 + e^{\eta_i})^2}.$$

Além disso, temos que

$$U_{\phi}(\boldsymbol{\beta}, \phi) = \text{tr}(\text{diag}\{u_1, \dots, u_n\}),$$

sendo $\text{tr}(\cdot)$ denotando o traço de uma matriz e

$$u_i = \log(-\log(y_i)) - \psi(\phi) - \log\left(\frac{\mu_i^{1/\phi}}{\alpha_i}\right) - \frac{1}{\phi} \alpha_i \log(\mu_i) \left[1 + \frac{\alpha_i \log(y_i)}{\phi \mu_i^{1/\phi}}\right],$$

onde $\psi(\phi) = d \log[\Gamma(\phi)]/d\phi$ é a função digama.

O estimador de máxima verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \phi)^{\top}$ é obtido numericamente através de algoritmo de otimização não linear (Newton-Raphson ou Escore de Fisher) ou algum algoritmo quasi-Newton (BFGS - Broyden-Fletcher-Goldfarb-Shanno), pois não é possível resolver analiticamente o sistema de equações $U(\boldsymbol{\beta}, \phi) = \mathbf{0}$.

A matriz de informação de Fisher é dada por

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

com

$$K_{\boldsymbol{\beta}\boldsymbol{\beta}} = X^\top W_{\boldsymbol{\beta}\boldsymbol{\beta}} X, \quad K_{\boldsymbol{\beta}\phi} = K_{\phi\boldsymbol{\beta}} = X^\top W_{\boldsymbol{\beta}\phi} \mathbf{1} \quad \text{e} \quad K_{\phi\phi} = \mathbf{1}^\top W_{\phi\phi} \mathbf{1},$$

em que

$$\begin{aligned} W_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \frac{1}{\phi} \text{diag} \left\{ \left[\frac{\alpha_i h'(\eta_i)}{\mu_i^{1/\phi+1}} \right]^2 \right\}, \\ W_{\boldsymbol{\beta}\phi} &= -\frac{1}{\phi} \text{diag} \left\{ \frac{\alpha_i}{\mu_i^{1/\phi+1}} \left[\frac{\alpha_i \log(\mu_i)}{\mu_i^{1/\phi}} + 1 \right] h'(\eta_i) \right\} \quad \text{e} \\ W_{\phi\phi} &= \text{diag} \left\{ \psi'(\phi) + \frac{2 \alpha_i \log(\mu_i)}{\phi^2 \mu_i^{1/\phi}} + \frac{1}{\phi^3} \left[\frac{\alpha_i \log(\mu_i)}{\mu_i^{1/\phi}} \right]^2 \right\} \end{aligned}$$

são matrizes diagonais de ordem $n \times n$, $i = 1, 2, \dots, n$. Além disso, temos que $\mathbf{1}$ é um vetor n -dimensional de uns e $\psi'(\phi) = d^2 \log[\Gamma(\phi)]/d\phi^2$ é a função trigama.

Vale ressaltar que, o vetor escore e a matriz de informação de Fisher são apresentados em Mousa et al. (2016) e Rocha (2020). No Apêndice A, mostramos os detalhes dos cálculos destas quantidades.

Capítulo 3

Inferência estatística via *bootstrap*

Em Estatística, temos dois procedimentos inferenciais importantes: estimação de parâmetros e testes de hipóteses. Neste capítulo, vamos apresentar uma das partes inéditas deste trabalho, que é o desenvolvimento de estudos de simulações (Seção 3.3) acerca destes procedimentos, usando reamostragem *bootstrap* (EFRON, 1979), nos modelos de regressão Gama Unitária.

3.1 Correção de viés

Viés de um estimador é a diferença entre o valor médio do estimador estatístico e o valor do parâmetro que se pretende estimar. Um estimador não-viesado é uma propriedade que assegura que, em média, o estimador do parâmetro desconhecido é correto. Assim, uma boa propriedade dos estimadores é serem não-viesados. Em muitas situações, quando a amostra é pequena ou até mesmo de tamanho moderado, os estimadores de máxima verossimilhança (EMVs) são viesados. Em geral, o viés do estimador de máxima verossimilhança é de ordem $O(n^{-1})$, onde n é o tamanho da amostra.

A correção de viés, em muitas situações, é necessária. Bartlett (1953) apresentou uma expressão para o viés de ordem n^{-1} do estimador de máxima verossimilhança no caso uniparamétrico. Cox e Snell (1968) chegaram a uma expressão geral para o viés de ordem n^{-1} dos estimadores de máxima verossimilhança nos casos uniparamétrico e multiparamétrico. Muitos estudos são baseados nos resultados de Cox e Snell (1968). Por exemplo, Cook et al. (1986) encontraram o viés dos EMVs para um modelo normal (não-linear); Cordeiro et al. (1997) con-

sideraram a distribuição beta e encontraram o viés dos EMVs baseado na abordagem de Cox e Snell (1968); Raydonal et al. (2006) derivaram expressões de forma fechada para correção de viés dos EMVs do modelo de regressão beta. Recentemente, Melo et al. (2020) também utilizaram a metodologia de Cox e Snell (1968) para obterem a correção de viés dos EMVs no modelo de regressão Dirichlet.

Para o uso da metodologia de Cox e Snell (1968) é necessário obter cumulantes que são resultantes do valor esperado das derivadas até quarta ordem da log-verossimilhança. Em alguns modelos, o cálculo destes cumulantes é tedioso. Uma alternativa ao método de Cox e Snell (1968) é o método de reamostragem *bootstrap*, proposto por Efron (1979). Este método substitui os cálculos analíticos de Cox e Snell (1968) por esforços computacionais. Usando a metodologia de Efron (1979) podemos obter estimadores corrigidos por meio de soluções computacionais. Neste trabalho, o método de reamostragem *bootstrap* foi empregado de forma paramétrica, quando há uma suposição sobre a distribuição de probabilidade dos dados. Cada amostra *bootstrap* é obtida da distribuição paramétrica que originou os dados.

Aqui, o objetivo é estimar um vetor de parâmetros de interesse $\boldsymbol{\theta} = \mathbf{t}(F)$ com base em uma amostra de tamanho n , $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$, vinda de uma distribuição F , considerada a amostra original (que representa a população). Para conseguir isso, é necessário calcular $\hat{\boldsymbol{\theta}} = \mathbf{s}(\mathbf{y})$. Essa estatística de interesse pode ser o erro padrão ou a estimativa do vetor de parâmetros $\boldsymbol{\theta}$. Uma amostra *bootstrap* $\mathbf{y}^* = (y_1^*, y_2^*, \dots, y_n^*)^\top$ é obtida por amostragem aleatória de tamanho n , com reposição, a partir dos dados originais \mathbf{y} . A distribuição *bootstrap* representa uma distribuição amostral desta estatística. Assim, os resultados só se tornam confiáveis se forem realizadas várias amostras *bootstrap* do mesmo tamanho n , com reposição e de forma aleatória. Para cada amostra *bootstrap* estima-se $\boldsymbol{\theta}$ através de B réplicas *bootstrap*, $\hat{\boldsymbol{\theta}}^* = \mathbf{s}(\mathbf{y}^*)$.

A estimativa *bootstrap* do erro padrão é o desvio padrão das réplicas *bootstrap*, sendo representado por

$$\hat{e}p_{boot}(\hat{\boldsymbol{\theta}}^*) = \sqrt{\left\{ \frac{1}{B-1} \sum_{b=1}^B [s(y_b^*) - \bar{s}(\mathbf{y}^*)]^2 \right\}},$$

em que $s(y_b^*)$ é o valor da estatística para cada amostra *bootstrap*, e

$$\bar{s}(\mathbf{y}^*) = \frac{1}{B} \sum_{b=1}^B s(y_b^*).$$

Essa técnica ainda permite que se encontre a estimativa *bootstrap* do viés do estimador $\hat{\boldsymbol{\theta}}$, que é dada por

$$\widehat{vies}_{boot}(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\theta}}_{boot}^* - \hat{\boldsymbol{\theta}},$$

em que

$$\hat{\boldsymbol{\theta}}_{boot}^* = \frac{1}{B} \sum_{b=1}^B \hat{\boldsymbol{\theta}}_b^*.$$

3.2 Testes de hipótese

Considere a partição do vetor $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top = (\boldsymbol{\psi}^\top, \boldsymbol{\omega}^\top)^\top$, com $\boldsymbol{\psi}^\top = (\beta_2, \beta_3, \dots, \beta_{q+1})$ representando um vetor q -dimensional de parâmetros de interesse e $\boldsymbol{\omega}^\top = (\beta_1, \beta_{q+2}, \dots, \beta_p, \phi)$ sendo o vetor $((p+1-q) \times 1)$ de parâmetros de perturbação. Note que, β_1 é o intercepto do modelo. O interesse consiste em testar

$$\mathcal{H}_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0 \quad \times \quad \mathcal{H}_1 : \boldsymbol{\psi} \neq \boldsymbol{\psi}_0,$$

em que $\boldsymbol{\psi}_0$ é um vetor de constantes de dimensão q .

O estimador de máxima verossimilhança restrito a hipótese nula \mathcal{H}_0 é denotado por $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\psi}_0^\top, \tilde{\boldsymbol{\omega}}^\top)^\top$, enquanto o irrestrito, denotado por $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}^\top, \hat{\boldsymbol{\omega}}^\top)^\top$. Em (2.5) definimos a função de log-verossimilhança por $\ell(\boldsymbol{\beta}, \phi)$, que pode ser denotada por $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\psi}, \boldsymbol{\omega})$. Além disso, usando a partição supracitada, podemos reescrever o vetor escore e a matriz de informação de Fisher definida no Capítulo 2, respectivamente, por

$$U(\boldsymbol{\theta}) = \begin{bmatrix} U_{\boldsymbol{\psi}}(\boldsymbol{\theta}) \\ U_{\boldsymbol{\omega}}(\boldsymbol{\theta}) \end{bmatrix} \quad \text{e} \quad K(\boldsymbol{\theta}) = \begin{bmatrix} K_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\theta}) & K_{\boldsymbol{\psi}\boldsymbol{\omega}}(\boldsymbol{\theta}) \\ K_{\boldsymbol{\omega}\boldsymbol{\psi}}(\boldsymbol{\theta}) & K_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\theta}) \end{bmatrix}.$$

A inversa da matriz de informação de Fisher é dada por

$$K(\boldsymbol{\theta})^{-1} = \begin{bmatrix} K^{\psi\psi}(\boldsymbol{\theta}) & K^{\psi\omega}(\boldsymbol{\theta}) \\ K^{\omega\psi}(\boldsymbol{\theta}) & K^{\omega\omega}(\boldsymbol{\theta}) \end{bmatrix}.$$

Os testes assintóticos (para grandes amostras) mais usuais na literatura são: razão de verossimilhanças (RV), escore (E), gradiente (G) e Wald (W). Estes testes têm distribuição, sob \mathcal{H}_0 , qui-quadrado com q graus de liberdade (χ_q^2), onde q é o número de restrições impostas pela hipótese nula. As estatísticas destes testes são dadas, respectivamente, por:

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

$$E = U_{\psi}(\tilde{\boldsymbol{\theta}})^{\top} K^{\psi\psi}(\tilde{\boldsymbol{\theta}}) U_{\psi}(\tilde{\boldsymbol{\theta}}),$$

$$G = U_{\psi}(\tilde{\boldsymbol{\theta}})^{\top} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0),$$

$$W = (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)^{\top} \left[K^{\psi\psi}(\hat{\boldsymbol{\theta}}) \right]^{-1} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0).$$

Vale ressaltar que, sob \mathcal{H}_0 , estas quatro estatísticas são assintoticamente equivalentes. Quando a amostra é pequena ou até mesmo de tamanho moderado, a convergência para a distribuição χ_q^2 destas estatísticas pode não ser boa e os testes não serão confiáveis (GUEDES et al., 2020; CORDEIRO e FERRARI, 1991; VARGAS et al., 2014; SEVERINI, 2000). Uma forma de contornar este problema é usar amostragem *bootstrap* paramétrica (EFRON, 1979), cujo objetivo é conseguir uma aproximação para a distribuição da estatística de teste sob \mathcal{H}_0 , a partir das reamostras *bootstrap*. Sob a hipótese nula, geramos um grande número de amostras *bootstrap* e calculamos as estatísticas de teste para cada uma delas. Todas as estatísticas de teste de *bootstrap* são usadas para estimar a distribuição das estatísticas de teste, sob a hipótese nula, e a partir daí obtemos um valor crítico para os testes (CRIBARI-NETO e QUEIROZ, 2014).

Segundo Cribari-Neto e Queiroz (2014), o algoritmo de *bootstrap* pode ser descrito como:

1. Calcule a estatística de teste de interesse $\tau(\mathbf{y})$ usando a amostra observada \mathbf{y} . Aqui $\tau(\mathbf{y})$ será RV , E , G e W ;
2. Substitua os parâmetros do modelo pelas estimativas de $\boldsymbol{\theta}$ sob \mathcal{H}_0 , obtidos com base na amostra \mathbf{y} e gere a amostra *bootstrap* \mathbf{y}^* ;
3. Use \mathbf{y}^* para estimar o modelo e calcule a estatística $\tau^* = \tau(\mathbf{y}^*)$;
4. Repita os passos 2 e 3 um número B grande de vezes;
5. Com base na distribuição empírica das B realizações da estatística τ^* (obtidas nos passos 2 a 4), calcule o quantil de interesse $\tau_{1-\alpha}^*$, onde α é o nível de nominal do teste;
6. Com a estatística $\tau(\mathbf{y})$, obtida no passo 1, e o valor crítico de *bootstrap* ($\tau_{1-\alpha}^*$), obtido no passo 5, realize o teste.

O quantil $\tau_{1-\alpha}^*$ é o que define a região de rejeição, de modo que, se a estatística for maior que o quantil estimado ($\tau(\mathbf{y}) > \tau_{1-\alpha}^*$), a hipótese nula é rejeitada. A análise também pode ser feita pelo valor- p aproximado por *bootstrap*:

$$\hat{\alpha}^* = \frac{\#\{\tau_b^* > \tau\}}{B},$$

sendo que τ_b^* é a estatística de teste calculada a partir da b -ésima amostra de *bootstrap*, \mathbf{y}_b^* , com $b = 1, 2, \dots, B$. Neste caso, se $\hat{\alpha}^*$ for menor que um dado α , a hipótese nula é rejeitada.

3.3 Resultados numéricos

Realizamos um estudo de simulação com 5.000 (cinco mil) réplicas de Monte Carlo e 1.000 (mil) amostras de *bootstrap* para avaliar os desempenhos dos testes da razão de verossimilhanças (RV), escore (E), gradiente (G) e Wald (W), as versões corrigidas são denotadas por *bootstrap*, RV_{boot} , E_{boot} , G_{boot} e W_{boot} , respectivamente, segundo a probabilidade do erro tipo I, ou seja, a probabilidade de rejeitar a hipótese nula quando a mesma é verdadeira, para níveis nominais dos testes pré-estabelecidos. Adicionalmente, avaliamos os comportamentos dos estimadores de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta}$. Para isso, obtemos as

estimativas dos vieses e as estimativas do erro quadrático médio (EQM) dos estimadores $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{boot}^*$.

As simulações foram realizadas utilizando o software R (R Core Team, 2021), baseadas no modelo de regressão Gama Unitária (2.4) para os tamanhos de amostra $n = 20, 30$ e 50 . Os valores das covariáveis foram obtidos de uma distribuição uniforme $U(-1/2, 1/2)$ e utilizamos a função de ligação logito, ou seja, $g(\mu) = \mu/(1 - \mu)$. Consideramos $p = 3$ e 4 covariáveis e testamos $q = 2$ e 3 parâmetros. Definimos os valores dos parâmetros como $\phi = 10$ e $\boldsymbol{\beta} = (1, \psi_0, 1, \dots, 1)^\top$. Os níveis nominais considerados foram $\alpha = 1\%, 5\%$ e 10% . O interesse neste estudo reside em testar $\mathcal{H}_0 : \boldsymbol{\psi} = \mathbf{0}$ contra a hipótese alternativa $\mathcal{H}_1 : \boldsymbol{\psi} \neq \mathbf{0}$, em que $\mathbf{0}$ é um vetor de zeros de dimensão $q \times 1$.

Nas Tabelas 3.1 e 3.2 apresentamos as estimativas de máxima verossimilhança (EMV), as estimativas dos vieses e as estimativas do EQM dos estimadores $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_{boot}^*$. Observamos que, na maioria dos casos, os vieses (em módulo) e o EQM são menores quando consideramos a correção pelo viés *bootstrap*. Isso fica mais evidente para o estimador de ϕ . Como esperado, a medida que o tamanho da amostra cresce os vieses e o EQM diminuem. Observamos ainda que, quando aumentamos o número de covariáveis no modelo, os vieses e o EQM também aumentam. E esse aumento é menor quando usamos *bootstrap*, principalmente para o parâmetro ϕ .

As taxas de rejeição da hipótese nula dos testes RV , E , G e W , e de suas versões corrigidas por *bootstrap*, RV_{boot} , E_{boot} , G_{boot} e W_{boot} , respectivamente, para os diferentes tamanhos amostrais n , números de variáveis explicativas p , número de parâmetros a serem testados q e níveis nominais α , podem ser observadas nas Tabelas 3.3 e 3.4. Vale salientar que, os valores das entradas são dadas em porcentagens. Observamos nestas tabelas que quando usamos amostragem *bootstrap*, as taxas de rejeição se aproximam de α , em pequenas amostras. Isto também ocorre quando a amostra cresce. Por exemplo, na Tabela 3.3 quando $n = 20$, $\alpha = 1\%$, $p = 3$, para o teste de Wald temos taxas $4,4\%$ (W) e $0,8\%$ (W_{boot}); para o teste Escore temos $0,3\%$ (E) e $1,0\%$ (E_{boot}). Além disso, à medida que o número de covariáveis aumenta, as taxas de rejeição ficam mais distantes dos respectivos α , mas quando usamos amostragem *bootstrap*, elas permanecem constantes e próximas dos níveis nominais usados.

Tabela 3.1: Estimativas do parâmetros: $q = 2$, $p = 3, 4$ e diferentes valores de n ; $\beta_1 = 1$, $\beta_2 = \beta_3 = 0$ e $\phi = 10$.

p	n	Parâmetros	$\hat{\theta}$			$\hat{\theta}_{boot}^*$		
			EMV	Viés	EQM	EMV	Viés	EQM
3	20	β_1	1,0019	0,0019	0,0068	0,9978	-0,0022	0,0068
		β_2	0,0005	0,0005	0,0763	0,0003	0,0003	0,0760
		β_3	0,0011	0,0011	0,0763	-0,0009	-0,0009	0,0909
		ϕ	13,2221	3,2221	35,2104	8,8999	-1,1001	12,3239
	30	β_1	1,0018	0,0018	0,0050	0,9987	-0,0013	0,0050
		β_2	0,0009	0,0009	0,0452	0,0001	0,0001	0,0450
		β_3	0,0004	0,0004	0,0671	-0,0025	-0,0025	0,0668
		ϕ	11,9604	1,9604	15,5831	6,6166	-0,3834	7,6772
	50	β_1	1,0009	0,0009	0,0029	0,9990	0,0011	0,0355
		β_2	0,0011	0,0011	0,0356	0,0011	0,0011	0,0355
		β_3	-0,0020	-0,0020	0,0423	-0,0032	-0,0032	0,0423
		ϕ	11,1016	1,1016	6,6154	9,8967	-0,1032	4,2932
4	20	β_1	1,0036	0,0036	0,0070	0,9985	-0,0015	0,0069
		β_2	-0,0026	-0,0026	0,0781	-0,0006	-0,0006	0,7775
		β_3	0,0000	0,0000	0,0935	-0,0018	-0,0018	0,0928
		β_4	1,0145	0,0145	0,0761	1,0089	0,0089	0,0754
		ϕ	14,1670	4,1670	48,7444	8,2223	-1,7777	13,7567
	30	β_1	1,0027	0,0027	0,0051	0,9989	-0,0011	0,0050
		β_2	-0,0003	-0,0003	0,0457	-0,0005	-0,0005	0,0455
		β_3	0,0001	0,0001	0,0687	-0,0032	-0,0032	0,0683
		β_4	1,0096	0,0096	0,0603	1,0069	0,0069	0,0600
		ϕ	12,4470	2,4470	19,2189	9,4008	-0,5992	7,8345
	50	β_1	1,0017	0,0017	0,0029	0,9993	-0,0007	0,0029
		β_2	-0,0002	-0,0002	0,0362	0,0003	0,0003	0,0361
β_3		-0,0011	-0,0011	0,4420	-0,0024	-0,0024	0,0440	
β_4		1,0100	0,0100	0,0335	1,0075	0,0075	0,0333	
ϕ		11,3532	1,3532	7,6326	9,8398	-0,1602	4,3674	

Tabela 3.2: Estimativas do parâmetros: $q = 3$, $p = 4$ e diferentes valores de n , $\beta_1 = 1$, $\beta_2 = \beta_3 = \beta_4 = 0$ e $\phi = 10$.

n	Parâmetros	$\hat{\theta}$			$\hat{\theta}_{boot}^*$		
		EMV	Viés	EQM	EMV	Viés	EQM
20	β_1	1,0038	0,0038	0,0070	0,9986	-0,0014	0,0069
	β_2	-0,0026	-0,0026	0,0772	-0,0009	-0,0009	0,0766
	β_3	0,0003	0,0003	0,0923	-0,0016	-0,0016	0,0916
	β_4	0,0090	0,0090	0,0756	0,0075	0,0075	0,0750
	ϕ	14,1661	4,1661	48,7361	8,2133	-1,7867	13,7776
30	β_1	1,0029	0,0029	0,0050	0,9990	-0,0010	0,0050
	β_2	-0,0003	-0,0003	0,0455	-0,0004	-0,0004	0,0454
	β_3	0,0002	0,0002	0,0677	-0,0028	-0,0028	0,0673
	β_4	0,0058	0,0058	0,0591	0,0062	0,0062	0,0589
	ϕ	12,4455	2,4455	19,2003	9,4001	-0,5999	7,8375
50	β_1	1,0018	0,0018	0,0029	0,9994	-0,0006	0,0029
	β_2	-0,0000	-0,0000	0,0361	0,0004	0,0004	0,0360
	β_3	-0,0009	-0,0010	0,0431	-0,0021	-0,0021	0,0429
	β_4	0,0077	0,0077	0,0332	0,0072	0,0072	0,0331
	ϕ	11,3532	1,3532	7,6358	9,8407	-0,1593	4,3677

Tabela 3.3: Taxas de rejeição nula: $q = 2$, $p = 3, 4$ e diferentes valores de n .

p	Estatística \ n	$\alpha = 1\%$			$\alpha = 5\%$			$\alpha = 10\%$			
		20	30	50	20	30	50	20	30	50	
3	RV	1,8	1,6	1,5	8,0	6,5	6,4	14,5	13,0	11,6	
	E	0,3	0,8	0,7	5,2	4,7	5,2	10,5	10,8	10,6	
	G	0,5	0,8	0,8	5,2	4,9	5,4	11,1	10,9	10,7	
	W	4,4	2,8	2,1	10,9	8,8	7,6	17,5	15,2	12,8	
	RV_{boot}	0,8	1,1	1,2	5,5	4,9	5,5	10,0	10,4	10,5	
	E_{boot}	1,0	1,3	1,1	5,4	4,8	5,4	9,9	10,4	10,6	
	G_{boot}	0,9	1,1	1,2	5,5	4,9	5,5	10,1	10,4	10,5	
	W_{boot}	0,8	1,1	1,3	5,6	5,1	5,5	10,3	10,5	10,6	
	4	RV	2,5	2,0	1,5	9,4	7,5	6,7	16,4	14,2	12,1
		E	0,6	1,0	0,9	6,0	5,3	5,5	12,3	12,0	10,8
G		0,6	0,9	0,9	6,1	5,6	5,7	12,8	12,2	11,0	
W		5,5	3,2	2,4	12,9	9,7	7,9	20,0	16,2	13,2	
RV_{boot}		0,8	1,2	1,2	5,4	5,0	5,2	10,5	10,5	10,2	
E_{boot}		0,9	1,2	1,1	5,3	5,0	5,2	10,2	10,4	10,4	
G_{boot}		0,8	1,2	1,2	5,4	5,0	5,2	10,5	10,5	10,2	
W_{boot}		0,8	1,3	1,2	5,6	5,1	5,2	10,3	10,3	10,5	

Tabela 3.4: Taxas de rejeição nula: $q = 3$, $p = 4$ e diferentes valores de n .

Estatística \ n	$\alpha = 1\%$			$\alpha = 5\%$			$\alpha = 10\%$		
	20	30	50	20	30	50	20	30	50
RV	2,4	1,7	1,4	9,5	7,4	6,8	16,6	14,1	12,6
E	0,3	0,6	0,8	4,3	4,3	5,1	10,5	10,0	10,6
G	0,3	0,6	0,6	4,6	4,5	5,2	10,9	10,2	11,0
W	6,4	3,8	2,7	15,0	10,7	8,8	21,6	18,0	14,6
RV_{boot}	1,2	1,0	1,0	5,0	4,8	5,3	10,5	9,8	10,8
E_{boot}	1,2	1,0	1,2	5,1	4,8	5,6	10,0	10,1	10,6
G_{boot}	1,2	1,1	1,1	5,1	4,8	5,6	10,5	9,8	10,9
W_{boot}	1,3	1,0	1,1	5,2	5,0	5,6	10,8	10,2	10,7

Capítulo 4

Aplicação

4.1 Descrição dos dados

Neste capítulo, apresentamos uma aplicação a dados de uma área que vem crescendo muito no Brasil nos últimos anos: a geração de energia através de sistemas fotovoltaicos. É comprovado que a microgeração de energia, que inclui a geração através de sistemas fotovoltaicos, tem impacto direto na geração distribuída, uma vez que pode reduzir custos, melhorar a confiabilidade e reduzir o impacto ambiental que é causado por usinas de grande porte. Além disso, a implementação desta fonte de energia alternativa colabora muito com a geração de empregos, este setor gerou mais de 86 mil empregos em 2020 (ABSOLAR, 2021b). É importante ressaltar a importância de fontes alternativas de energia devido a inevitáveis épocas de seca, momentos estes impactam diretamente na geração de energia por meio de usinas hidroelétricas, que são as maiores responsáveis por esta atividade no Brasil atualmente, como foi noticiado pelo próprio site do governo brasileiro (www.gov.br) em janeiro de 2020, ao citar que 63,8% de sua matriz energética vem da geração através desta fonte. Nessas situações, o apelo se voltaria a fontes de energia não renováveis, se não fosse o surgimento das alternativas, tais como a fotovoltaica. Então, levando em conta os fatores atrativos ligados a este setor, como diversificação da matriz energética; redução de custos e de perdas em todo o processo da geração à distribuição de energia via geração concentrada de energia; economia de água; diminuição do uso de fontes não renováveis; geração de empregos; fácil instalação e conexão na rede de distribuição de

baixa tensão; surge a oportunidade de estudar os impactos da microgeração distribuída, que vem crescendo cada vez mais – teve um aumento de 39% no Brasil até agosto de 2021 (ABSOLAR, 2021c). Este setor demonstra, também, um enorme potencial futuro, já que a geração distribuída solar pode gerar 173 bilhões de reais em redução de custos aos consumidores até 2050 (ABSOLAR, 2021a), e também que este setor deve atrair até 70 bilhões de reais em 10 anos (ABSOLAR, 2020). Outro fator que mostra o tamanho do potencial de crescimento desta área é a redução nos preços das peças. Com a maior popularidade da utilização de sistemas fotovoltaicos e o aumento da oferta do produto, os materiais utilizados nas instalações (como os módulos solares, inversores, estruturas de suporte, baterias e controladores de carga) já apresentam valores menores, e isso tende a melhorar ainda mais com o passar do tempo. A reportagem da ISTOÉ (ISTOÉ, 2021), em janeiro de 2022, relatou que um sistema fotovoltaico residencial custava, em média, R\$ 35.080,00 em junho de 2016; já em junho de 2021, o valor caiu para R\$ 19.520,00, o que representa uma redução de cerca de 40%. Além disso, *fintechs* e bancos passaram a financiar o sistema solar.

Diante do exposto, a variável dependente usada na aplicação, deste capítulo, é o percentual da potência instalada da geração distribuída por estado brasileiro. Por geração distribuída, compreende-se casas e empresas de pequeno e médio porte, ou seja, geração de energia em locais onde se aplica baixa tensão. Os dados da variável dependente foram obtidos no site da Associação Brasileira de Energia Solar Fotovoltaica (ABSOLAR, 2021c). Por uma questão de falta de dados para a variável supracitada, em algumas regiões, foram selecionadas 9 capitais. As variáveis independentes consideradas foram: a tarifa da energia de baixa tensão, o *payback*, o IFDM e a taxa de ocupação destas capitais.

A tarifa aplicada ao consumidor de baixa tensão (tensões até 1000 V que se aplicam ao público que utiliza uma carga de até 75 KW) é um fator importante inclusive no estudo de viabilidade para verificar se é benéfico fazer a instalação de um sistema fotovoltaico. Esta questão, sobretudo a partir de 2018, vem sendo mais estudada no que diz respeito à mudança na forma de cobrança. Esta mudança pode impactar diretamente na porcentagem de consumidores do sistema fotovoltaico, uma vez que impacta diretamente no tempo de retorno do projeto, inclusive. Sobre os valores das tarifas, é importante ressaltar que, independente do tipo de consumidor, na conta de luz são levados em conta os seguintes fatores: geração,

transporte de energia até o consumidor (que envolve transmissão e distribuição) e tributos e encargos. Cada concessionária e estado possui esses custos diferentes, uma vez que estas tarifas levam em conta características de cada área de concessão da distribuidora, ainda que regulamentadas pela ANEEL. Além do custo com a aquisição da energia elétrica, a tarifa aplicada ao consumidor leva em conta mais 4 fatores: custos relativos ao uso do sistema de distribuição; custos relativos ao uso do sistema de transmissão; perdas técnicas (energia dissipada pelos condutores quando há passagem de corrente) e não técnicas (problemas com medidores); e encargos diversos e tributos. Os tributos incidentes na tarifa são: Imposto sobre Circulação de Mercadorias e Serviços (ICMS), Programa de Integração Social (PIS) e a Contribuição para o Financiamento de Seguridade Social (Cofins).

Outra variável independente considerada é o payback, que é fundamental no projeto de investimento de alguém que deseja investir na geração de energia por meio da fonte renovável supracitada. Esta é uma forma de definir o tempo de retorno sobre um investimento inicial. A partir deste indicador, o investidor consegue obter mais um parâmetro para definir se é viável ou não entrar em determinado projeto. Portanto, é fundamental verificar se esta variável impacta percentual da potência instalada da geração distribuída por estado brasileiro, já que, teoricamente, o consumidor deve levar em conta este tempo de retorno para verificar se vai instalar este sistema de fato. Segundo Claudio Loureiro da Array, que trabalha em uma empresa americana de energia solar, em entrevista à Revista ISTOÉ (ISTOÉ, 2021), o payback médio, nestes casos, era de 5 anos, e hoje pode ser calculado em 4 anos, com a bandeira vermelha. Este tempo curto de retorno de investimento para consumidores, principalmente residenciais, pode ser mais um atrativo em prol do crescimento deste setor.

O IFDM é o Índice Firjan de Desenvolvimento Municipal. Neste estudo, foram levados em conta os dados desta variável referente ao ano de 2016. Criado em 2008, ele é feito, exclusivamente, com base em estatísticas públicas oficiais, disponibilizadas pelos ministérios do Trabalho, Educação e Saúde (FIRJAN, 2018). MELO (2017) mostrou uma relação positiva entre o IFDM e a implementação de parques eólicos – que também são fontes de energia renováveis - na Bahia, o que desperta o interesse da relação deste com a variável dependente aqui estudada. Este índice é uma referência para avaliar o desenvolvimento socioeconômico, e acompanha quatro áreas: emprego, renda, educação e saúde. A leitura deste dado funciona da

seguinte forma: ele varia entre 0 e 1, e quanto mais próximo de 1, maior do desenvolvimento da localidade estudada. Outro ponto interessante a respeito deste indicador é que ele mostra se a melhoria é em decorrência da adoção de políticas do próprio local e se é apenas reflexo dos demais municípios. Se espera que este índice faça parte do modelo por mostrar a realidade socioeconômica de cada capital, o que impacta no poder de informação e de capital, por exemplo, dos indivíduos. No caso desta variável, foram estabelecidos 4 conceitos para o IFDM: os locais com valores entre 0 e 0,4 apresentam um baixo estágio de desenvolvimento; entre 0,4 e 0,6 possuem desenvolvimento regular; entre 0,6 e 0,8; desenvolvimento moderado; e, por fim, entre 0,8 e 1, alto estágio de desenvolvimento.

A última variável independente considerada diz respeito à taxa de ocupação de pessoas nas localidades estudadas. É importante ressaltar o que isso significa. Na metodologia utilizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), uma pessoa ocupada é aquela que, exerce atividade profissional (formal ou informal, remunerada ou não) durante no mínimo uma hora completa na semana de referência da pesquisa, veja IBGE (2020). É importante considerar, inicialmente, esta variável neste estudo pelo fato de que, para o consumidor usufruir deste tipo de geração de energia, é necessário investimento, e para poder investir, é necessário que se tenha capital. Portanto, há a expectativa de que isto tenha um impacto positivo sobre a variável explicada, que é o percentual da potência instalada da geração distribuída por estado brasileiro.

Os dados são apresentados na Tabela 4.1.

4.2 Modelagem

O modelo geral considerado inicialmente é o modelo de regressão Gama-Unitária, dado por

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i}, \quad i = 1, 2, \dots, n, \quad (4.1)$$

em que x_{1i} é a taxa de baixa tensão, x_{2i} é o IFDM, x_{3i} é a taxa de ocupação e x_{4i} é o payback. As estimativas dos parâmetros são $\hat{\beta}_0 = 28,214$, $\hat{\beta}_1 = -0,0314$, $\hat{\beta}_2 = -1,938$, $\hat{\beta}_3 = 5,164$, $\hat{\beta}_4 = -2,396$ e $\phi = 24,098$. E as respectivas estimativas corrigidas, via *bootstrap* são: $\hat{\beta}_0^* = 49,894$, $\hat{\beta}_1^* = -0,054$, $\hat{\beta}_2^* = -5,747$, $\hat{\beta}_3^* = 9,557$, $\hat{\beta}_4^* = -4,121$ e $\phi^* = 44,268$.

Tabela 4.1: Dados usados na aplicação.

Capitais	Potência instalada (%)	Tarifa Baixa Tensão (R\$ x MWh)	Pay Back Anos	IFDM IFDM	População ocupada (%)
Belo Horizonte	0,6	708,29	5,67	0,822	58,7
Brasília	1,1	552,86	6,36	0,780	46,2
Campo Grande	0,7	430,22	8,28	0,815	33,3
Cuiabá	1,3	522,01	6,69	0,827	44,6
Fortaleza	0,9	633,69	5,18	0,745	31,8
Goiânia	0,8	639,18	5,50	0,817	43,6
Manaus	0,6	593,75	6,24	0,693	23,7
Rio de Janeiro	0,8	725,13	5,20	0,789	37,1
Teresina	0,9	586,49	5,66	0,828	35,1

Fonte: ABSOLAR (2021c), Índice COMERC (2016), IFDM (2018)

Inicialmente, testamos a hipótese de $\mathcal{H}_0 : (\beta_1, \beta_2) = (0, 0)^\top$ contra $\mathcal{H}_1 : (\beta_1, \beta_2) \neq (0, 0)^\top$, foram obtidos os seguintes valores: teste de Razão de Verossimilhança, Gradiente, Escore e Wald apresentaram valores- p , sem correção, respectivamente, de 1,000; 0,000; 0,219 e 0,003. Os mesmos testes, ainda respectivamente, apresentaram valores- p , com correção, de 0,058; 0,020; 0,067 e 0,066. Isso significa que, a um nível de significância de 5%, os testes de Razão de Verossimilhança (sem e com correção), Gradiente (com correção), Escore (sem e com correção) e Wald (com correção), não apresentam indícios para rejeitar a hipótese nula, enquanto apenas nos testes Gradiente e Wald (ambos sem correção), há rejeição da hipótese nula. Como este teste analisou a relevância dos dois primeiros parâmetros (Taxa de Baixa Tensão e IFDM), eles devem ser retirados do modelo e, então, temos o modelo reduzido:

$$\text{logit}(\mu_i) = \beta_0 + \beta_3 x_{3i} + \beta_4 x_{4i}, \quad i = 1, 2, \dots, n. \quad (4.2)$$

Neste caso, As estimativas dos parâmetros são $\hat{\beta}_0 = -4,938$, $\hat{\beta}_3 = 0,548$, $\hat{\beta}_4 = 0,007$ e $\phi = 112,371$. E as respectivas estimativas corrigidas, via *bootstrap* são: $\hat{\beta}_0^* = -9,863$, $\hat{\beta}_3^* = 1,097$, $\hat{\beta}_4^* = 0,013$ e $\phi^* = 224,370$.

Consideramos agora o teste $\mathcal{H}_0 : (\beta_3, \beta_4) = (0, 0)^\top$ contra $\mathcal{H}_1 : (\beta_3, \beta_4) \neq (0, 0)^\top$, o resultado foi: teste de Razão de Verossimilhança, Gradiente, Escore e Wald apresentaram

valores- p , sem correção, respectivamente, de 1,000; 0,945; 0,963 e 0,938. Todos os testes corrigidos apresentaram valores- p de 0,002. Ou seja, a um nível de significância de 5%, sem correção, todos os testes não rejeitariam a hipótese nula e descartariam as variáveis restantes. Porém, com a correção, todos eles rejeitam, o que significa que os parâmetros não devem ser descartados.

Outro fator indicativo é o *AIC* (do português, Critério de Informação de Akaike), que ajuda a escolher, dentre modelos, o que minimiza a divergência Kulback-Leiber, que está relacionada à informação perdida, uma vez que o modelo se trata de uma aproximação da realidade, e não dela entendida na sua totalidade, com todos os dados à disposição (BOZDOGAN, 1987). Desta forma, o modelo com o menor valor é considerado, por este quesito, o melhor. O valor para o primeiro modelo apresentado neste estudo foi de $AIC = -49,979$, enquanto o do segundo foi de $AIC = -75,384$. Logo, pela explicação dada, há mais um indicativo de que o último se mostra, de fato, um modelo mais confiável.

Em suma, se conclui disto que as variáveis Taxa de Baixa Tensão e IFDM não são significativas para a geração de energia nas capitais brasileiras através de sistemas fotovoltaicos; e com relação à Taxa de Ocupação e o payback, a dedução é o oposto (as variáveis se mostram significativas).

Capítulo 5

Conclusões

Existem em muitas situações práticas, dados de taxas e proporções. Nestes casos, alguns modelos são usados na literatura, como por exemplo, o modelo de regressão Gama Unitária. Este trabalho se mostra um importante contribuinte para a pesquisa por comprovar que a aplicação da reamostragem via *bootstrap*, de fato, impacta positivamente no modelo abordado. Ou seja, nos direciona para uma interpretação correta dos impactos de variáveis dependentes sobre a variável resposta. Foram realizados estudos através de simulações de Monte Carlo, via *bootstrap*, para obtenção de estimadores de máxima verossimilhança e testes de hipóteses corrigidos para tamanho de amostras pequeno. Estes estudos numéricos foram feitos usando o software R (R Core Team, 2021). Avaliamos numericamente os desempenhos dos estimadores e dos testes originais e corrigidos e verificamos que as versões corrigidas apresentam comportamento mais confiáveis em amostras de tamanhos pequeno do que os estimadores e testes usuais.

Outra contribuição importante desta dissertação foi uma aplicação a dados reais, em que a variável dependente foi o percentual da potência instalada da geração distribuída por estado brasileiro. Inicialmente, quatro variáveis independentes foram consideradas no modelo, a saber: Taxa de Baixa Tensão, IFDM, Taxa de ocupação e *payback*. Após o uso dos testes corrigidos, propostos aqui, as duas primeiras variáveis não foram significativas. Assim, no modelo final, temos a taxa de ocupação e o *payback* explicando o percentual da potência instalada de energia fotovoltaica.

Enfim, neste trabalho, observamos que o método *bootstrap* para correção de viés de estimadores e testes de hipóteses tem um desempenho muito bom, quando consideramos o modelo de regressão Gama Unitário na modelagem dos dados.

Apêndice A

Vetor escore e matriz de informação de Fisher

Neste apêndice, apresentamos, matricialmente, o vetor escore, a matriz de informação observada e a matriz de informação de Fisher (esperada) mostradas no Capítulo 2. Estas quantidades foram obtidas em Mousa et al. (2016) e Rocha (2020).

A.1 Densidade da Gama Unitária

Vimos, no Capítulo 2, que $W \sim \text{Gama}(\alpha, \phi)$. Para obter a distribuição Gama Unitária a partir da distribuição Gama, Rocha (2020) usa a transformação $Y = e^{-W}$, obtendo $Y \sim \text{UG}(\alpha, \phi)$. Abaixo temos a função distribuição acumulada da variável Y ,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(e^{-W} \leq y) \\ &= P(-W \leq \log y) = P(W \geq -\log y) \\ &= 1 - P(W < -\log y) \\ &= 1 - F_W(-\log y). \end{aligned}$$

Derivando com respeito a y , obtém-se

$$f(y) = -g(-\log y) \left(-\frac{1}{y} \right) = \frac{g(-\log y)}{y}.$$

Mas

$$g(w; \alpha, \phi) = \frac{\alpha^\phi}{\Gamma(\phi)} e^{-\alpha w} w^{\phi-1}.$$

Portanto,

$$\begin{aligned} f(y) &= \frac{\alpha^\phi}{\Gamma(\phi)} e^{-\alpha(-\log y)} (-\log y)^{\phi-1} \frac{1}{y} \\ f(y) &= \frac{\alpha^\phi}{\Gamma(\phi)} y^{\alpha-1} (-\log y)^{\phi-1}. \end{aligned}$$

Para obter a f.d.p. da Gama Unitária em termos de μ e ϕ , vamos calcular, primeiramente, o valor da Esperança $E(Y)$. Temos

$$\begin{aligned} E(Y) &= \int_0^1 y f(y) dy \\ &= \int_0^1 y \frac{\alpha^\phi}{\Gamma(\phi)} y^{\alpha-1} (-\log y)^{\phi-1} dy \\ &= \frac{\alpha^\phi}{\Gamma(\phi)} \int_0^1 y^\alpha (-\log y)^{\phi-1} dy. \end{aligned}$$

Se usarmos a f.d.p. da distribuição $UG(\alpha + 1, \phi)$, obtemos que

$$\int_0^1 y^\alpha (-\log y)^{\phi-1} dy = \frac{\Gamma(\phi)}{(\alpha + 1)^\phi}.$$

Logo,

$$E(Y) = \frac{\alpha^\phi}{\Gamma(\phi)} \frac{\Gamma(\phi)}{(\alpha + 1)^\phi}.$$

Portanto,

$$E(Y) = \left(\frac{\alpha}{\alpha + 1} \right)^\phi.$$

Considerando que $E(Y) = \mu$, temos

$$\mu = \left(\frac{\alpha}{\alpha + 1} \right)^\phi.$$

Logo,

$$\alpha = \frac{\mu^{1/\phi}}{1 - \mu^{1/\phi}}. \quad (\text{A.1})$$

Desta forma, a f.d.p. da Gama Unitária em termos de μ e ϕ é dada por

$$f(y; \mu, \phi) = \frac{1}{\Gamma(\phi)} \left[\frac{\mu^{1/\phi}}{1 - \mu^{1/\phi}} \right]^\phi y^{[\mu^{1/\phi}/(1-\mu^{1/\phi})]^{-1}} (-\log y)^{\phi-1}.$$

A.2 Vetor escore

A função de verossimilhança é representada por

$$\begin{aligned} L(\mu_i, \phi) &= \prod_{i=1}^n f(y_i; \alpha_i, \phi) \\ &= \prod_{i=1}^n \left\{ \frac{\alpha_i^\phi}{\Gamma(\phi)} y_i^{\alpha_i-1} (-\log y_i)^{\phi-1} \right\} \\ &= \left[\prod_{i=1}^n \Gamma(\phi)^{-1} \right] \left[\prod_{i=1}^n \alpha_i^\phi \right] \left[\prod_{i=1}^n y_i \right]^{\alpha_i-1} \left[\prod_{i=1}^n (-\log y_i) \right]^{\phi-1}. \end{aligned}$$

O logaritmo da função de verossimilhança é dado por

$$\begin{aligned} l(\mu_i, \phi) &= \log[L(\mu_i, \phi)] \\ &= \log \left\{ \left[\prod_{i=1}^n \Gamma(\phi)^{-1} \right] \left[\prod_{i=1}^n \alpha_i^\phi \right] \left[\prod_{i=1}^n y_i \right]^{\alpha_i-1} \left[\prod_{i=1}^n (-\log y_i) \right]^{\phi-1} \right\} \\ &= \sum_{i=1}^n \phi \log \alpha_i - \sum_{i=1}^n \log[\Gamma(\phi)] + \sum_{i=1}^n (\alpha_i - 1) \log y_i + \sum_{i=1}^n (\phi - 1) \log(-\log y_i) \\ &= \sum_{i=1}^n \{ \phi \log \alpha_i - \log[\Gamma(\phi)] + (\alpha_i - 1) \log y_i + (\phi - 1) \log(-\log y_i) \}. \end{aligned}$$

Portanto, a equação acima pode ser escrita como

$$\ell(\boldsymbol{\beta}, \phi) = \sum_{i=1}^n \ell_i(\mu_i, \phi), \quad (\text{A.2})$$

com

$$\ell_i(\mu_i, \phi) = \phi \log \alpha_i - \log[\Gamma(\phi)] + (\alpha_i - 1) \log y_i + (\phi - 1) \log(-\log y_i),$$

sendo que $\alpha_i = \mu_i^{1/\phi} / (1 - \mu_i^{1/\phi})$. A primeira derivada de (A.2) com relação a β_j é

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i, \phi)}{\partial \beta_j},$$

em que

$$\frac{\partial \ell_i(\mu_i, \phi)}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \left\{ \phi \log \alpha_i - \log[\Gamma(\phi)] + (\alpha_i - 1) \log y_i + (\phi - 1) \log(-\log y_i) \right\}. \quad (\text{A.3})$$

Pode-se escrever, ainda, a partir de (2.4), que $\mu_i = g^{-1}(\eta_i)$. Fazendo $g^{-1}(\cdot) = h(\cdot)$, conclui-se que

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

Daí,

$$\frac{\partial \mu_i}{\partial \beta_j} = h'(\eta_i) \frac{\partial \eta_i}{\partial \beta_j},$$

em que $h'(\eta_i) = \partial h(\eta_i) / \partial \beta_j$. Além disso,

$$\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial (X_j^\top \boldsymbol{\beta})}{\partial \beta_j} = x_{ij}.$$

Portanto,

$$\frac{\partial \mu_i}{\partial \beta_j} = h'(\eta_i) x_{ij}.$$

Além disso,

$$\frac{\partial \alpha_i}{\partial \beta_j} = \frac{\partial \alpha_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}.$$

Mas,

$$\begin{aligned}\frac{\partial \alpha_i}{\partial \mu_i} &= \frac{\partial}{\partial \mu_i} \left\{ \frac{\mu_i^{1/\phi}}{1 - \mu_i^{1/\phi}} \right\} \\ &= \frac{(1/\phi)\mu_i^{(1/\phi)-1}(1 - \mu_i^{1/\phi}) - \mu_i^{1/\phi}[-(1/\phi)\mu_i^{(1/\phi)-1}]}{[1 - \mu_i^{1/\phi}]^2} \\ \frac{\partial \alpha_i}{\partial \mu_i} &= \frac{\mu_i^{(1/\phi)-1}}{\phi [1 - \mu_i^{1/\phi}]^2}.\end{aligned}$$

Sabemos que

$$\alpha_i = \frac{\mu_i^{1/\phi}}{1 - \mu_i^{1/\phi}}.$$

Então,

$$1 - \mu_i^{1/\phi} = \frac{\mu_i^{1/\phi}}{\alpha_i}.$$

Logo,

$$\begin{aligned}\frac{\partial \alpha_i}{\partial \mu_i} &= \frac{\mu_i^{1/\phi}}{1 - \mu_i^{1/\phi}} \frac{\mu_i^{-1}}{\phi(1 - \mu_i^{1/\phi})} \\ &= \alpha_i \frac{\mu_i^{-1}}{\phi \frac{\mu_i^{1/\phi}}{\alpha_i}} \\ \frac{\partial \alpha_i}{\partial \mu_i} &= \frac{\alpha_i^2}{\phi \mu_i^{1/\phi+1}}.\end{aligned}$$

Assim,

$$\frac{\partial \alpha_i}{\partial \beta_j} = \frac{\alpha_i^2}{\phi \mu_i^{1/\phi+1}} h'(\eta_j) x_{ij}.$$

Voltando na equação (A.3):

$$\begin{aligned}
\frac{\partial \ell_i(\mu_i, \phi)}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \left\{ \phi \log \alpha_i - \log(\Gamma(\phi) + \alpha_i - 1) \log y_i + (\phi - 1) \log(-\log y_i) \right\} \\
&= \left\{ \frac{\phi}{\alpha} \frac{\partial \alpha_i}{\partial \beta_j} + \log y_i \frac{\partial \alpha_i}{\partial \beta_j} \right\} \\
&= \left[\frac{\alpha_i}{\mu_i^{1/\phi+1}} + \frac{\alpha_i^2 \log y_i}{\phi \mu_i^{1/\phi+1}} \right] h'(\eta_j) x_{ij}.
\end{aligned}$$

Portanto,

$$\frac{\partial \ell_i(\mu_i, \phi)}{\partial \beta_j} = (\mu_i^* + y_i^*) h'(\eta_j) x_{ij}, \quad (\text{A.4})$$

com

$$\mu_i^* = \frac{\alpha_i}{\mu_i^{1/\phi+1}} \quad \text{e} \quad y_i^* = \frac{\alpha_i^2 \log y_i}{\phi \mu_i^{1/\phi+1}}. \quad (\text{A.5})$$

Faça

$$\begin{aligned}
X &= \begin{bmatrix} \mathbf{X}_1^\top \\ \mathbf{X}_2^\top \\ \vdots \\ \mathbf{X}_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \\
T &= \begin{bmatrix} h'(\eta_1) & 0 & \cdots & 0 \\ 0 & h'(\eta_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & h'(\eta_p) \end{bmatrix} \\
\text{e } \mathbf{s} &= \begin{bmatrix} \mu_1^* + y_1^* \\ \mu_2^* + y_2^* \\ \vdots \\ \mu_n^* + y_n^* \end{bmatrix}.
\end{aligned}$$

Daí,

$$\begin{aligned} U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi) &= \frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta}} = X^{\top} T \mathbf{s} \\ &= \begin{bmatrix} \sum_{i=1}^n (\mu_i^* + y_i^*) h'(\eta_i) x_{i1} \\ \sum_{i=1}^n (\mu_i^* + y_i^*) h'(\eta_i) x_{i2} \\ \vdots \\ \sum_{i=1}^n (\mu_i^* + y_i^*) h'(\eta_i) x_{ip} \end{bmatrix} = \begin{bmatrix} \frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_1} \\ \frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_2} \\ \vdots \\ \frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_p} \end{bmatrix}. \end{aligned}$$

Portanto,

$$U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi) = X^{\top} T \mathbf{s}.$$

Agora, é preciso calcular

$$U_{\phi}(\boldsymbol{\beta}, \phi) = \frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \phi}.$$

Mas,

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \phi} = \sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\beta}, \phi)}{\partial \phi},$$

com

$$\begin{aligned} \frac{\partial \ell_i(\boldsymbol{\beta}, \phi)}{\partial \phi} &= \frac{\partial}{\partial \phi} \left\{ \phi \log \alpha_i - \log(\Gamma(\phi)) + (\alpha_i - 1) \log y_i + (\phi - 1) \log(-\log y_i) \right\} \\ &= \log \alpha_i + \phi \frac{1}{\alpha_i} \frac{\partial \alpha_i}{\partial \phi} - \Psi(\phi) + \frac{\partial \alpha_i}{\partial \phi} \log y_i + \log(-\log y_i), \end{aligned}$$

em que

$$\psi(\phi) = \frac{\partial}{\partial \phi} \log \Gamma(\phi).$$

Agora, é preciso encontrar

$$\frac{\partial \alpha_i}{\partial \phi} = \frac{\partial}{\partial \phi} \left[\frac{\mu_i^{1/\phi}}{1 - \mu_i^{1/\phi}} \right].$$

Usando a propriedade da função exponencial, temos que

$$\mu_i^{1/\phi} = e^{(1/\phi) \log \mu_i}.$$

Então,

$$\begin{aligned}\frac{d\mu_i^{1/\phi}}{d\phi} &= e^{(1/\phi)\log\mu_i} \left(-\frac{1}{\phi^2}\right) \log\mu_i \\ &= -\frac{\mu_i^{1/\phi}}{\phi^2} \log\mu_i.\end{aligned}$$

Portanto,

$$\begin{aligned}\frac{\partial\alpha_i}{\partial\phi} &= \frac{-\frac{\mu_i^{1/\phi}}{\phi^2} \log\mu_i (1 - \mu_i^{1/\phi}) - \mu_i^{1/\phi} \left[\frac{\mu_i^{1/\phi}}{\phi^2} \log\mu_i\right]}{\left[1 - \mu_i^{1/\phi}\right]^2} \\ &= \alpha_i \left[-\frac{\log\mu_i}{\phi^2(1 - \mu_i^{1/\phi})}\right].\end{aligned}$$

A partir de (A.1),

$$\frac{1}{1 - \mu_i^{1/\phi}} = \frac{\alpha_i}{\mu_i^{1/\phi}},$$

então,

$$\begin{aligned}\frac{\partial\alpha_i}{\partial\phi} &= \alpha_i \left(-\frac{\log\mu_i}{\phi^2}\right) \frac{1}{1 - \mu_i^{1/\phi}} \\ &= -\frac{\alpha_i^2 \log\mu_i}{\phi^2 \mu_i^{1/\phi}}\end{aligned}$$

Daí,

$$\frac{\partial\ell_i(\boldsymbol{\beta}, \phi)}{\partial\phi} = \log\alpha_i + \frac{\phi}{\alpha_i} \left[-\frac{\alpha_i^2 \log\mu_i}{\phi^2 \mu_i^{1/\phi}}\right] - \Psi(\phi) + \log y_i \left[-\frac{\alpha_i^2 \log\mu_i}{\phi^2 \mu_i^{1/\phi}}\right] + \log(-\log y_i) \quad (\text{A.6})$$

$$= \log(-\log y_i) - \Psi(\phi) - \log\left(\frac{\mu_i^{1/\phi}}{\alpha_i}\right) - \frac{1}{\phi} \alpha_i \log\mu_i \left[1 + \frac{\alpha_i \log y_i}{\phi \mu_i^{1/\phi}}\right] \quad (\text{A.7})$$

Portanto,

$$U_\phi(\boldsymbol{\beta}, \phi) = \sum_{i=1}^n u_i,$$

em que

$$u_i = \log(-\log y_i) - \Psi(\phi) - \log\left(\frac{\mu_i^{1/\phi}}{\alpha_i}\right) - \frac{1}{\phi}\alpha_i \log \mu_i \left[1 + \frac{\alpha_i \log y_i}{\phi \mu_i^{1/\phi}}\right].$$

O vetor score assume a forma

$$U((\boldsymbol{\beta}, \phi)) = \begin{bmatrix} U_\beta(\boldsymbol{\beta}, \phi) \\ U_\phi(\boldsymbol{\beta}, \phi) \end{bmatrix}.$$

A.3 Matriz de informação de Fisher

A matriz de informação de Fisher é dada por

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

em que

$$K_{\beta\beta} = X^\top W_{\beta\beta} X, \quad K_{\beta\phi} = K_{\phi\beta} = X^\top W_{\beta\phi} \mathbf{1} \quad \text{e} \quad K_{\phi\phi} = \mathbf{1}^\top W_{\phi\phi} \mathbf{1}.$$

De (A.4), temos

$$\frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \beta_j \partial \beta_k} = \left[\frac{\partial \mu_i^*}{\partial \beta_j} + \frac{\partial y_i^*}{\partial \beta_j} \right] h'(\eta_i) x_{ik} + (\mu_i^* + y_i^*) h''(\eta_i) \frac{\partial \eta_i}{\partial \beta_j} x_{ik}.$$

De (A.5), conclui-se que

$$\begin{aligned} \frac{\partial \mu_i^*}{\partial \beta_j} &= \frac{1}{\left[\mu_i^{1/\phi+1}\right]^2} \left\{ \left(\frac{\partial \alpha_i}{\partial \beta_j}\right) \mu_i^{1/\phi+1} - \alpha_i \left(\frac{1}{\phi} + 1\right) \mu_i^{1/\phi} \frac{\partial \mu_i}{\partial \beta_j} \right\} \\ &= \frac{\alpha_i h'(\eta_i) x_{ij}}{\phi \left[\mu_i^{1/\phi+1}\right]^2} \left(\alpha_i - \mu_i^{1/\phi} - \phi \mu_i^{1/\phi} \right), \end{aligned}$$

e também que

$$\begin{aligned} \frac{\partial y_i^*}{\partial \beta_j} &= \frac{1}{[\mu_i^{1/\phi+1}]^2} \left\{ 2\alpha_i \frac{\partial \alpha_i}{\partial \beta_j} \mu_i^{1/\phi+1} - \alpha_i^2 \left(\frac{1}{\phi} + 1 \right) \mu_i^{1/\phi} \frac{\partial \mu_i}{\partial \beta_j} \right\} \frac{\log y_i}{\phi} \\ &= \frac{\alpha_i^2 h'(\eta_i) x_{ij} \log y_i}{\phi^2 [\mu_i^{1/\phi+1}]^2} \left(2\alpha_i - \mu_i^{1/\phi} - \phi \mu_i^{1/\phi} \right), \end{aligned}$$

Portanto,

$$\begin{aligned} -\frac{\partial^2 \ell_i(\mu_t, \phi)}{\partial \beta_j \partial \beta_k} &= -\frac{\alpha_i [h'(\eta_i)]^2 x_{ij} x_{ik}}{\phi^2 [\mu_i^{1/\phi+1}]^2} \left[\phi \alpha_i - \phi \mu_i^{1/\phi} - \phi^2 \mu_i^{1/\phi} + 2\alpha_i^2 \log y_i - \alpha_i \log y_i \mu_i^{1/\phi} \right. \\ &\quad \left. - \alpha_i \log y_i \phi \mu_i^{1/\phi} \right] - (\mu_i^* + y_i^*) h''(\eta_i) x_{ij} x_{ik}. \end{aligned}$$

Sabemos que

$$\begin{aligned} E \left[\frac{\partial \ell_i(\mu_i, \phi)}{\partial \beta_k} \right] &= 0 \\ E [(\mu_i^* + y_i^*) h'(\eta_i) x_{ik}] &= 0 \\ E (\mu_i^* + y_i^*) h'(\eta_i) x_{ik} &= 0 \\ E (\mu_i^* + y_i^*) &= 0. \end{aligned}$$

Além disso,

$$\begin{aligned} E (\mu_i^*) + E (y_i^*) &= 0 \\ E (y_i^*) &= -\mu_i^* \frac{\alpha_i^2}{\phi \mu_i^{1/\phi+1}} \\ E[\log y_i] &= -\frac{\alpha_i}{\mu_i^{1/\phi+1}} \\ E[\log y_i] &= -\frac{\phi}{\alpha_i}. \end{aligned}$$

Daí,

$$\begin{aligned}
 E \left[-\frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \beta_j \partial \beta_k} \right] &= -\frac{\alpha_i [h'(\eta_i)]^2 x_{ij} x_{ik}}{\phi^2 [\mu_i^{1/\phi+1}]^2} [\phi \alpha_i - \phi \mu_i^{1/\phi} - \phi^2 \mu_i^{1/\phi} + 2\alpha_i^2 E(\log y_i) - \alpha_i E(\log y_i) \mu_i^{1/\phi} \\
 &\quad - \alpha_i E(\log y_i) \phi \mu_i^{1/\phi}] - E(\mu_i^* + y_i^*) h''(\eta_i) x_{ij} x_{ik} \\
 &= -\frac{\alpha_i^2 [h'(\eta_i)]^2 x_{ij} x_{ik}}{\phi [\mu_i^{1/\phi+1}]^2}.
 \end{aligned}$$

Logo,

$$\begin{aligned}
 K_{\beta_j \beta_k} &= E \left[-\frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \beta_j \partial \beta_k} \right] \\
 &= \sum_{i=1}^n x_{ij} \frac{\alpha_i^2 [h'(\eta_i)]^2}{\phi [\mu_i^{1/\phi+1}]^2} x_{ik} \\
 &= \sum_{i=1}^n x_{ij} \frac{\alpha_i^2}{\phi} [h'(\eta_i)]^2 x_{ik}
 \end{aligned}$$

em que $\alpha_i = \alpha_i / (\mu_i^{1/\phi+1})$. Daí,

$$\begin{aligned}
 K_{\beta\beta} &= \begin{bmatrix} K_{\beta_1\beta_1} & K_{\beta_1\beta_2} & \cdots & K_{\beta_1\beta_p} \\ K_{\beta_2\beta_1} & K_{\beta_2\beta_2} & \cdots & K_{\beta_2\beta_p} \\ \vdots & \vdots & \vdots & \vdots \\ K_{\beta_p\beta_1} & K_{\beta_p\beta_2} & \cdots & K_{\beta_p\beta_p} \end{bmatrix} = X^\top W_{\beta\beta} X \\
 &= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \times \begin{bmatrix} (\alpha_1^2/\phi)h'(\eta_1)^2 & 0 & \cdots & 0 \\ 0 & (\alpha_2^2/\phi)h'(\eta_2)^2 & \cdots & 0 \\ \vdots & 0 & \vdots & \vdots \\ 0 & 0 & \cdots & (\alpha_n^2/\phi)h'(\eta_n)^2 \end{bmatrix} \\
 &\times \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \\
 &= \begin{bmatrix} \sum_{i=1}^n x_{i1}(\alpha_i^2/\phi)h'(\eta_i)^2 x_{i1} & \cdots & \sum_{i=1}^n x_{i1}(\alpha_i^2/\phi)h'(\eta_i)^2 x_{ip} \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^n x_{ip}(\alpha_i^2/\phi)h'(\eta_i)^2 x_{i1} & \cdots & \sum_{i=1}^n x_{ip}(\alpha_i^2/\phi)h'(\eta_i)^2 x_{ip} \end{bmatrix}
 \end{aligned}$$

Portanto,

$$K_{\beta\beta} = X^\top W_{\beta\beta} X.$$

Agora, de (A.6), temos

$$\begin{aligned}
\frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \phi^2} &= \frac{\partial}{\partial \phi} \left\{ \log(-\log y_t) - \Psi(\phi) - \log\left(\frac{\mu_i^{1/\phi}}{\alpha_i}\right) - \frac{1}{\phi} \alpha_i \log \mu_i \left[1 + \frac{\alpha_i \log y_i}{\phi \mu_i^{1/\phi}} \right] \right\} \\
&= -\frac{\partial \Psi(\phi)}{\partial \phi} - \left(\frac{-1}{\phi^2}\right) \log \mu_i + \frac{1}{\alpha_i} \frac{\partial \alpha_i}{\partial \phi} - \frac{\partial}{\partial \phi} \left[\frac{\alpha_i}{\phi} \log \mu_t + \frac{\alpha_i^2}{\phi^2 \mu_i^{1/\phi}} \log(y_i) \log(\mu_i) \right] \\
&= -\Psi'(\phi) + \frac{\log(\mu_i)}{\phi^2} - \frac{\alpha_i \log(\mu_i)}{\phi^2 \mu_i^{1/\phi}} + \frac{\alpha_i^2 [\log(\mu_i)]^2}{\phi^3 \mu_i^{1/\phi}} + \frac{\alpha_i \log(\mu_i)}{\phi^2} \\
&+ \frac{2\alpha_i^3}{\phi^4 [\mu_i^{1/\phi}]^2} \log(y_i) [\log(\mu_i)]^2 + \frac{2\alpha_i^2 \log(\mu_i)}{\phi^3 \mu_i^{1/\phi}} \log(y_i) - \frac{\alpha_i^2 [\log(\mu_i)]^2}{\phi^4 \mu_i^{1/\phi}} \log(y_i).
\end{aligned}$$

Sabemos que $E[\log(y_t)] = -\phi/\alpha_t$. Então,

$$\frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \phi^2} = \Psi'(\phi) + \frac{2c_i}{\phi^2} + \frac{c_i^2}{\phi^3} + \frac{\log(\mu_i)}{\phi^2} \left(-(\alpha_i + 1) + \frac{\alpha_i}{\mu_i^{1/\phi}} \right),$$

em que, $c_i = [\alpha_i \log(\mu_i)]/(\mu_i^{1/\phi})$. Mas $\alpha_i + 1 = \alpha_i/(\mu_i^{1/\phi})$. Portanto,

$$\frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \phi^2} = \Psi'(\phi) + \frac{2c_i}{\phi^2} + \frac{c_i^2}{\phi^3}.$$

Assim, temos

$$\begin{aligned}
 K_{\phi\phi} &= E \left[-\frac{\partial^2 \ell(\boldsymbol{\beta}, \phi)}{\partial \phi^2} \right] \\
 &= \sum_{i=1}^n E \left[-\frac{\partial^2 \ell_i(\boldsymbol{\beta}, \phi)}{\partial \phi^2} \right] \\
 &= \sum_{i=1}^n E \left[\Psi'(\phi) + \frac{2c_i}{\phi^2} + \frac{c_i^2}{\phi^3} \right] \\
 &= \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \Psi'(\phi) + 2c_1/\phi^2 + c_1^2/\phi^3 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \Psi'(\phi) + 2c_n/\phi^2 + c_n^2/\phi^3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \\
 &= \mathbf{1}^\top W_{\phi\phi} \mathbf{1}.
 \end{aligned}$$

Agora, ainda de (A.6), temos

$$\begin{aligned}
 \frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \beta_j \partial \phi} &= \frac{\partial}{\partial \beta_j} \left\{ \log(-\log y_i) - \Psi(\phi) - \log \left(\frac{\mu_i^{1/\phi}}{\alpha_i} \right) - \frac{1}{\phi} \alpha_i \log \mu_i \left[1 + \frac{\alpha_i \log y_i}{\phi \mu_i^{1/\phi}} \right] \right\} \\
 &= - \left\{ \frac{1}{\phi} \frac{1}{\mu_i} \frac{\partial \mu_i}{\partial \beta_j} - \frac{1}{\alpha_i} \frac{\partial \alpha_i}{\partial \beta_j} + \frac{1}{\phi} \left[\frac{\partial \alpha_i}{\partial \beta_j} \log(\mu_i) + \frac{\alpha_i}{\mu_i} \frac{\partial \mu_i}{\partial \beta_j} + \frac{\log(y_i)}{\phi^2 [\mu_i^{1/\phi}]^2} \left(\left[2\alpha_i \frac{\partial \alpha_i}{\partial \beta_j} \log(\mu_i) \right. \right. \right. \right. \\
 &\quad \left. \left. \left. + \frac{\alpha_i^2}{\mu_i} \frac{\partial \mu_i}{\partial \beta_j} \right] \phi \mu_i^{1/\phi} - \alpha_i^2 \log(\mu_i) \mu_i^{1/\phi-1} \frac{\partial \mu_i}{\partial \beta_j} \right) \right] \right\}.
 \end{aligned}$$

Mas $\partial \mu_i / \partial \beta_j = h'(\eta_i) x_{ij}$. Então,

$$\begin{aligned}
 \frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \beta_j \partial \phi} &= -h'(\eta_i) x_{ij} \left\{ \frac{1}{\phi \mu_i} - \frac{1}{\alpha_i} \frac{\alpha_i^2}{\phi \mu_i^{1/\phi+1}} + \frac{1}{\phi} \left[\frac{\alpha_i^2}{\phi \mu_i^{1/\phi+1}} \log(\mu_i) + \frac{\alpha_i}{\mu_i} \right. \right. \\
 &\quad \left. \left. + \frac{\log(y_i)}{\phi^2 [\mu_i^{1/\phi}]^2} \left(\left[2\alpha_i \frac{\alpha_i^2}{\phi \mu_i^{1/\phi+1}} \log(\mu_i) + \frac{\alpha_i^2}{\mu_i} \right] \phi \mu_i^{1/\phi} - \alpha_i^2 \log(\mu_i) \mu_i^{1/\phi-1} \right) \right] \right\}.
 \end{aligned}$$

Sabemos que $E[\log(y_i)] = -\phi/\alpha_i$. Então,

$$E \left[-\frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \beta_j \partial \phi} \right] = \frac{h'(\eta_i) x_{ij}}{\phi} \left\{ \frac{1}{\mu_i} - \frac{\alpha_i}{\mu_i^{1/\phi+1}} + \frac{\alpha_i^2 \log(\mu_i)}{\phi \mu_i^{1/\phi+1}} + \frac{\alpha_i}{\mu_i} - \frac{2\alpha_i^2 \log(\mu_i)}{\phi \mu_i^{1/\phi+1} \mu_i^{1/\phi}} \right. \\ \left. - \frac{\alpha_i}{\mu_i^{1/\phi+1}} + \frac{\alpha_i \log(\mu_i)}{\phi \mu_i^{1/\phi+1}} \right\}.$$

Mas $\alpha_i + 1 = \alpha_i/(\mu_i^{1/\phi})$. Portanto,

$$E \left[-\frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \beta_j \partial \phi} \right] = \frac{h'(\eta_i) x_{ij}}{\phi} \left\{ \frac{1}{\mu_i} \frac{\alpha_i}{\mu_i^{1/\phi}} - \frac{2\alpha_i}{\mu_i^{1/\phi+1}} + \frac{\alpha_i \log(\mu_i)}{\phi \mu_i^{1/\phi+1}} \frac{\alpha_i}{\mu_i^{1/\phi}} - \frac{2\alpha_i^2 \log(\mu_i)}{\phi \mu_i^{1/\phi+1} \mu_i^{1/\phi}} \right\} \\ = \frac{h'(\eta_i) x_{ij}}{\phi \mu_i^{1/\phi+1}} \left\{ -\alpha_i - \frac{\alpha_i^2 \log(\mu_i)}{\phi \mu_i^{1/\phi}} \right\} \\ = -\frac{\alpha_i}{\mu_i^{1/\phi+1}} \frac{1}{\phi} \left\{ \frac{1}{\phi} \frac{\alpha_i \log(\mu_i)}{\mu_i^{1/\phi}} + 1 \right\} h'(\eta_i) x_{ij}.$$

Fazendo $a_i = \alpha_i/(\mu_i^{1/\phi+1})$, $c_i = \alpha_i \log(\mu_i)/(\mu_i^{1/\phi})$ e $b_i = -a_i$, temos

$$E \left[-\frac{\partial^2 \ell_i(\mu_i, \phi)}{\partial \beta_j \partial \phi} \right] = \frac{b_i}{\phi} \left(\frac{c_i}{\phi} + 1 \right) h'(\eta_i) x_{ij}.$$

Portanto,

$$\begin{aligned}
K_{\beta\phi} &= K_{\phi\beta} = \begin{bmatrix} K_{\beta_1\phi} & K_{\beta_1\phi} & \cdots & K_{\beta_1\phi} \\ K_{\beta_2\phi} & K_{\beta_2\phi} & \cdots & K_{\beta_2\phi} \\ \vdots & \vdots & \vdots & \vdots \\ K_{\beta_p\phi} & K_{\beta_p\phi} & \cdots & K_{\beta_p\phi} \end{bmatrix} = \\
&= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \\
&\times \begin{bmatrix} (b_1/\phi)(c_1/\phi + 1)h'(\eta_1) & 0 & \cdots & 0 \\ 0 & (b_2/\phi)(c_2/\phi + 1)h'(\eta_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & (b_n/\phi)(c_n/\phi + 1)h'(\eta_n) \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n (b_i/\phi)(c_i/\phi + 1)h'(\eta_i)x_{i1} & \cdots & \sum_{i=1}^n (b_i/\phi)(c_i/\phi + 1)h'(\eta_i)x_{ip} \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^n (b_i/\phi)(c_i/\phi + 1)h'(\eta_i)x_{i1} & \cdots & \sum_{i=1}^n (b_i/\phi)(c_i/\phi + 1)h'(\eta_i)x_{ip} \end{bmatrix} \\
&= X^\top W_{\beta\phi} \mathbf{1}.
\end{aligned}$$

Referências Bibliográficas

- [1] ABSOLAR (2020). Associação Brasileira de Energia Solar Fotovoltaica, <https://www.absolar.org.br/noticia/geracao-distribuida-de-energia-deve-atrair-ate-r70-bi-em-10-anos-preve-governo>
- [2] ABSOLAR (2021a). Associação Brasileira de Energia Solar Fotovoltaica, <https://www.absolar.org.br/noticia/geracao-distribuida-solar-pode-gerar-r-173-bilhoes-em-reducao-de-custos-aos-consumidores-ate-2050-aponta-absolar>, acessado em 25/08/2022.
- [3] ABSOLAR (2021b). Associação Brasileira de Energia Solar Fotovoltaica, <https://www.absolar.org.br/noticia/setor-fotovoltaico-gerou-mais-de-86-mil-empregos-no-brasil-em-2020>, acessado em 25/08/2022.
- [4] ABSOLAR (2021c). Associação Brasileira de Energia Solar Fotovoltaica, <https://www.absolar.org.br/mercado/infografico/>, acessado em 25/08/2022.
- [5] BARNDORFF-NIELSEN, O. E. (1991). Modified signed log likelihood ratio. *Biometrika*, v. 78, p. 557–563.
- [6] BARTLETT, M. S. (1937). Properties of sufficiency and statistical test. *In: Proceeding of the Royal Society A*, v. 160, p. 268-282.
- [7] CARRASCO, J. M.F.; FERRARI, S. L. P.; ARELLANO-VALLE, R. B. (2014). Errors-in-variables beta regression models. *Journal of Applied Statistics*, v.41, p. 1530-1547.

- [8] COOK, R.; TSAI, C.; WEI, B. (1986). Bias in nonlinear regression. *Biometrika*, v. 73, p. 615-623.
- [9] CORDEIRO, G. M.; FERRARI, S. L. P. (1991). A modified score test statistic having chi-squared distribution to order n^{-1} . *Biometrika*, v. 78, p. 573–582.
- [10] CORDEIRO, G. M.; ROCHA, E. C.; ROCHA, J. G. C.; CRIBARI-NETO, F. (1997). Bias-corrected maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, v. 58, p. 21-35.
- [11] CRIBARI-NETO, F.; QUEIROZ, M. P. F. (2014). On testing inference in beta regressions. *Journal of Statistical Computation and Simulation*, v. 84, p. 186-203.
- [12] EFRON, G. M. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, v. 7, p. 1-26.
- [13] EFRON, B.; TIBSHIRANI, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman and Hall.
- [14] ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. (2008). Influence diagnostics in beta regression. *Computational Statistics & Data Analysis*, v. 52, p. 4417-4431.
- [15] FIRJAN (2018). <https://www.firjan.com.br/ifdm/>, acessado em 25/08/2022.
- [16] GUEDES, A. C.; CRIBARI-NETO, F.; ESPINHEIRA, P. L. (2020). Modified likelihood ratio tests for unit gamma regressions. *Journal of Applied Statistics*, v.47, p. 1562-1586.
- [17] IBGE (2020). Instituto Brasileiro de Geografia e Estatística, https://biblioteca.ibge.gov.br/visualizacao/periodicos/2421/pnact_2020_2tri.pdf, acessado em 25/08/2022.
- [18] IFDM (2018), Índice FIRJAN Desenvolvimento Municipal, *Recorte Municipal-Abrangência Nacional*, Ano Base 2016, www.firjan.com.br/ifdm.
- [19] ISTOÉ (2021). Revista ISTOÉ, <https://istoe.com.br/producao-de-energia-solar-dobra-com-crise-hidrica/>, acessado em 25/08/2022.

- [20] LINDLEY, D. V. (1958). Fiducial Distributions and Bayes Theorem. *Journal of the Royal Statistical Society: Series B (Methodological)*, v.20, p.102-107.
- [21] MAZUCHELI, J., M.; MENEZES, A., F., B.; CHAKRABORTY, S. (2019). On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *Journal of Applied Statistics*, v.46, p.700-714.
- [22] MELO, J. L. (2017). Análise dos Impactos econômicos da implantação de parques eólicos no Estado da Bahia, Pontifícia Universidade Católica de São Paulo (PUC-SP), Mestrado em Administração, São Paulo, <https://tede2.pucsp.br/bitstream/handle/20466/2/Jefferson%20Lisboa%20Melo.pdf>, acessado em 25/08/2022.
- [23] MELO, T. F. N.; VARGAS, T. M. ; LEMONTE, A. J. ; MORENO-ARENAS, G. (2020). On improved estimation in multivariate Dirichlet regressions. *Communications in Statistics. Theory and Methods (Online)*, v. 49, p. 5765-5777.
- [24] MOUSA, A. M.; EL-SHEIKH, A. A.; ABDEL-FATTAH, M. A. A. (2016). Gamma Regression for Bounded Continuous Variables. *Advances and Applications in Statistics*, v. 49, 305–326.
- [25] NEYMAN, J.; e PEARSON, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, v. 20, 175–240.
- [26] OSPINA, R.; CRIBARI-NETO, F; VASCONCELOS, K. L. P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis*, v. 51, p. 960-981.
- [27] R CORE TEAM (2021). **R : A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- [28] RAO, C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, v. 44, p. 50-57.

- [29] SEVERINI, T. A. (2000). *Likelihood Methods in Statistics*. Oxford University Press.
- [30] SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis*, v.54, p. 348-366.
- [31] ROCHA, S. S. (2020). *Diagnóstico em modelos de regressão Gama Unitária*. Tese (Doutorado em Estatística) - Universidade Federal de Pernambuco, Recife.
- [32] ROCHA, S. S.; ESPINHEIRA, P. L.; CRIBARI-NETO, F. (2021). Residual and local influence analyses for unit gamma regressions. *Statistica Neerlandica*, v. 75, n. 2, p. 137–160.
- [33] SKOVGAARD, I. M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics*, v. 28, p. 3–32.
- [34] TERRELL, G. R. (2002). The Gradient Statistic. *Computing Science and Statistics*, v. 34, p. 206-215.
- [35] VARGAS, T. M.; FERRARI, S. L. P.; LEMONTE, A. J. (2014). Improved likelihood inference in generalized linear models. *Computational Statistics & Data Analysis*, v. 74, p. 110-124.
- [36] WALD, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society*, v. 54, p. 426-482.