



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Paula Tárcimam Gomes Santos

Modelo de Regressão GOLL- Σ Aplicado a Dados de Vegetação

Recife

2025

Paula Tércimam Gomes Santos

Modelo de Regressão GOLL- Σ Aplicado a Dados de Vegetação

Este trabalho foi apresentado ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Área de Concentração: Estatística Aplicada

Orientador (a): Prof. Dr. Abraão Nascimento

Coorientador (a): Prof. Dr. Jodavid Ferreira

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Santos, Paula Tárcimam Gomes.

Modelo de regressão GOLL-T aplicado a dados de vegetação /
Paula Tárcimam Gomes Santos. - Recife, 2025.
62f.: il.

Dissertação (Mestrado) - Universidade Federal de Pernambuco,
Centro de Ciências Exatas e da Natureza, Programa de Pós-
Graduação em Estatística, 2025.

Orientação: Abraão Nascimento.

Coorientação: Jodavid Ferreira.

Inclui referências e apêndices.

1. Distribuições de probabilidade; 2. GOLL-T; 3. Índices de
vegetação; 4. Modelos de regressão; 5. Sensoriamento remoto. I.
Nascimento, Abraão. II. Ferreira, Jodavid. III. Título.

UFPE-Biblioteca Central



UNIVERSIDADE FEDERAL DE PERNAMBUCO

Ata da defesa/apresentação do Trabalho de Conclusão de Curso de **Mestrado** do Programa de Pós-graduação em Estatística - CCEN da Universidade Federal de Pernambuco, no dia 29 de maio de 2025.

ATA Nº 237

Aos 29 dias de maio de 2025, às 14 horas, em sessão pública realizada de forma remota, teve início a defesa do Trabalho de Conclusão de Curso intitulada “UM NOVO MODELO DE REGRESSÃO GOLL-T COM APLICAÇÃO A DADOS DE VEGETAÇÃO” da mestranda PAULA TÁRCIMAM GOMES SANTOS, na área de concentração Estatística Aplicada, sob a orientação do Prof. Dr. Abraão David Costa do Nascimento e, coorientação do Prof. Dr. Jodavid de Araújo Ferreira. A Comissão Examinadora foi aprovada Ad referendum pelo coordenador do Programa de Pós-Graduação em Estatística, sendo composta pelos examinadores: Prof. Dr. Abraão David Costa do Nascimento (Presidente); Profa. Dra. Maria do Carmo Soares de Lima (Titular Interna); e Prof. Dr. Esdras Adriano Barbosa Dos Santos (Titular Externo). Após cumpridas as formalidades conduzidas pelo presidente da comissão, Prof. Dr. Abraão David Costa do Nascimento, a candidata ao grau de Mestra foi convidada a discorrer sobre o conteúdo do Trabalho de Conclusão de Curso. Concluída a explanação, a candidata foi arguida pela Comissão Examinadora que, em seguida, reuniu-se para deliberar e conceder, a mesma, a menção APROVADA. Para a obtenção do grau de Mestra em Estatística, a concluinte deverá ter atendido todas às demais exigências estabelecidas no Regimento Interno e Normativas Internas do Programa, nas Resoluções e Portarias dos Órgãos Deliberativos Superiores, assim como no Estatuto e no Regimento Geral da Universidade, observando os prazos e procedimentos vigentes nas normas.

Prof. Dr. Abraão David Costa do Nascimento
Presidente (Orientador), UFPE

Profa. Dra. Maria do Carmo Soares de Lima
Examinadora Interna à Instituição, UFPE

Prof. Dr. Esdras Adriano Barbosa Dos Santos
Examinador Externo à Instituição, UFS

PAULA TÁRCIMAM GOMES SANTOS
Mestranda

Dedico este trabalho, antes de tudo, a Deus, por ser meu sustento, meu refúgio e a razão de cada conquista. À minha mãe e ao meu pai, que oraram por mim com fé e esperança em cada etapa dessa caminhada. E, de forma muito especial, ao meu tio Luciano e à minha tia Rizo, que tanto amo e que foram presença constante, exemplo de carinho, apoio e luz nos momentos em que mais precisei.

Dedico também à minha querida avó Severina, que mesmo não estando mais entre nós, permaneceu viva em minhas lembranças e orações. Sua preocupação, suas palavras de incentivo e sua fé inabalável me acompanharam ao longo dessa jornada. Ela vibrou em cada vitória, chorou comigo nas dificuldades e sempre me cobriu com suas orações e amor. Sua memória é parte da minha força.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por conduzir meus passos até aqui, por me conceder o dom da sabedoria e da inteligência, por ser minha força nos dias de desânimo e por estar presente em cada momento desta caminhada.

Aos meus pais, por todo amor, apoio e ensinamentos que me sustentam. Aos meus tios Luciano e Rizo, minha sincera gratidão pela presença, incentivo e carinho constantes ao longo dessa trajetória.

Em memória da minha querida avó Severina, cuja presença permanece viva em mim. Com seu amor, orações e fé, acompanhou cada passo da minha caminhada. Vibrou com minhas conquistas, chorou nas minhas dores e sempre acreditou em mim. A ela, minha eterna gratidão e saudade.

De forma muito especial, agradeço aos meus orientadores, Professor Doutor Abraão Nascimento e Professor Doutor Jodavid Ferreira. Suas orientações firmes, sua escuta atenta e generosa e o carinho com que me acompanharam foram fundamentais para o amadurecimento deste trabalho e da minha formação como pesquisadora.

Estendo meu agradecimento a todos os professores do Departamento de Estatística da UFPE, em especial ao coordenador Professor Doutor Pablo Martins, que sempre se mostrou solícito e comprometido com nossa formação. Registro ainda minha gratidão aos Professores Doutores Maria do Carmo e Roberto Ferreira, por suas valiosas contribuições e incentivo constante.

Sou profundamente grata aos amigos que caminharam comigo nesta jornada, pelo apoio, pelas conversas e pelas partilhas. Em especial, ao Mestre Kleber Henrique e à Mestra Maria Mariana, cuja amizade e generosidade foram pilares importantes no meu percurso.

Minha gratidão também aos técnicos, secretários e prestadores de serviço da universidade, que, com dedicação e zelo, colaboram silenciosamente para o bom funcionamento das atividades acadêmicas. Sua contribuição é essencial e muitas vezes invisível, mas profundamente valiosa.

Por fim, meu reconhecimento à Universidade Federal de Pernambuco, instituição pública de excelência, essencial para o desenvolvimento científico e social do nosso país. Agradeço também à bolsa concedida pelo CNPq, que foi fundamental para meu crescimento acadêmico e para a realização desta dissertação.

RESUMO

A análise de índices de vegetação é fundamental para a produção agrária e a qualidade do meio ambiente. Dentre os índices possíveis de serem analisados, o NDVI (*Normalized Difference Vegetation Index*) e o EVI (*Enhanced Vegetation Index*) têm recebido muita atenção por sua relação fácil com os diferentes tipos de vegetação. Contudo, a análise desses índices requer o trabalho com distribuições tanto flexíveis como de suporte limitado. Para esse fim, esta dissertação propõe uma nova distribuição, denominada GOLL- $\mathfrak{T}$, construída a partir da família Geométrica Log-Logística Ímpar e reparametrizada em termos da média com suporte no intervalo $(0,2)$. A motivação para o desenvolvimento dessa distribuição está na necessidade de modelar o índice de vegetação NDVI modificado a partir do EVI, utilizado no monitoramento da vegetação via sensoriamento remoto. Com base nessa distribuição, é proposto um modelo de regressão que incorpora propriedades analíticas relevantes, como momentos, entropia de Shannon, funções score e matriz de informação de Fisher em forma matricial. Um estudo de simulação Monte Carlo avalia a performance dos estimadores de máxima verossimilhança em diferentes cenários, evidenciando boa performance conforme o aumento do tamanho amostral. A aplicação do modelo a dados reais obtidos via plataforma SATVeg mostra que a abordagem proposta é capaz de capturar relações não lineares entre variáveis ambientais de maneira mais eficaz do que modelos lineares tradicionais.

Palavras-chaves: Distribuições de probabilidade, GOLL- $\mathfrak{T}$, Índices de vegetação, Modelos de regressão, Sensoriamento remoto.

ABSTRACT

The analysis of vegetation indices is essential for agricultural production and environmental quality. Among the indices that can be examined, the NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index) have attracted considerable attention due to their straightforward relationship with different types of vegetation. However, analyzing these indices requires working with distributions that are both flexible and defined on a limited support. To address this, the present dissertation proposes a new distribution, called GOLL- $\mathfrak{T}$, constructed from the Odd Log-Logistic Geometric family and reparameterized in terms of the mean, with support on the interval $(0,2)$. The motivation behind developing this distribution lies in the need to model the modified NDVI, derived from EVI, used in vegetation monitoring through remote sensing. Based on this distribution, a regression model is proposed that incorporates relevant analytical properties such as moments, Shannon entropy, score functions, and the Fisher information matrix in matrix form. A Monte Carlo simulation study assesses the performance of maximum likelihood estimators under different scenarios, showing good performance as the sample size increases. The application of the model to real data obtained from the SATVeg platform demonstrates that the proposed approach can capture nonlinear relationships between environmental variables more effectively than traditional linear models.

Keywords: GOLL- $\mathfrak{T}$, Probability distributions, Regression models, Remote sensing, Vegetation indices.

LISTA DE FIGURAS

Figura 1 – Densidades da distribuição GOLL- $\mathcal{T}$ com diferentes valores do parâmetro μ .	19
Figura 2 – Variância	21
Figura 3 –	22
Figura 4 – Entropia de Shannon	23
Figura 5 – Mapa de Pernambuco com localização pontual (centroide) de cada município mais o arquipélago de Fernando de Noronha.	37
Figura 6 – Autocorrelação espacial com Índice de Moran Global e mapa de Pernambuco com informações do índice de Moran Local (LISA).	38
Figura 7 – Mapas de Pernambuco com informações de NDVI* e EVI	39
Figura 8 – Ajuste do modelos de regressão GOLL- $\mathcal{T}$ e modelo de regressão linear. . . .	41
Figura 9 – Análise dos Resíduos	42
Figura 10 – Mapas de Pernambuco com valores da entropia observada e predita para cada município	44

LISTA DE TABELAS

Tabela 1 – Resultados da simulação com diferentes tamanhos da amostra quando os parâmetros $\beta_0 = 1$ e $\beta_1 = -2$, e $\bar{\beta}$ é a média das estimativas e $\hat{b}_{\bar{\beta}}$ é o viés.	30
Tabela 2 – Resultados da simulação com diferentes tamanhos da amostra quando os parâmetros $\beta_0 = 1$ e $\beta_1 = 2$, e $\bar{\beta}$ é a média das estimativas e $\hat{b}_{\bar{\beta}}$ é o viés.	31
Tabela 3 – Resultados da simulação com diferentes tamanhos da amostra quando os parâmetros $\beta_0 = 1$, $\beta_1 = -2$ e $\beta_2 = -1$, e $\bar{\beta}$ é a média das estimativas e $\hat{b}_{\bar{\beta}}$ é o viés.	32
Tabela 4 – Resultados da simulação com diferentes tamanhos da amostra quando os parâmetros $\beta_0 = 1$, $\beta_1 = 2$ e $\beta_2 = 1$ e $\bar{\beta}$ é a média das estimativas e $\hat{b}_{\bar{\beta}}$ é o viés.	32
Tabela 5 – Estimação pontual, erro padrão, estatística de teste t e P-valor	40

LISTA DE SÍMBOLOS

γ Letra grega Gama

$\in$ Pertence

δ Delta

θ Teta

σ Sigma

μ Mi

SUMÁRIO

1	INTRODUÇÃO	13
1.1	MOTIVAÇÃO	13
1.2	CONTRIBUIÇÃO DESTA DISSERTAÇÃO	14
1.3	PLATAFORMA COMPUTACIONAL	14
1.3.1	Texto e computação	14
1.3.2	SATVeg - Sistema de Análise Temporal da Vegetação	14
1.3.2.1	<i>Dados Utilizados</i>	15
1.4	ORGANIZAÇÃO DA DISSERTAÇÃO	15
2	DISTRIBUIÇÃO GOLL-$\mathfrak{T}$	17
2.1	DEFININDO A FAMÍLIA GOLL- $\mathfrak{T}$	17
2.2	UMA NOVA DISTRIBUIÇÃO REPARAMETRIZADA (GOLL- $\mathfrak{T}$)	17
2.3	PROPRIEDADES DA GOLL- $\mathfrak{T}(\mu)$	19
2.3.1	Esperança e Variância	19
2.3.2	Assimetria e Curtose	21
2.4	ENTROPIA DE SHANNON PARA GOLL- $\mathfrak{T}(\mu)$	22
3	MODELO DE REGRESSÃO GOLL-$\mathfrak{T}$	24
3.1	REFERENCIAL TEÓRICO	24
3.2	REGRESSÃO GOLL- $\mathfrak{T}$	25
3.3	ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA	25
3.3.1	Log-Verossimilhança	25
3.3.2	Função <i>Escore</i>	26
3.3.3	Matriz Informação de Fisher	27
3.3.4	Método <i>Escore</i>-Fisher	28
3.3.5	Resíduo <i>Deviance</i>	29
4	ESTUDO DE SIMULAÇÃO E APLICAÇÃO EM DADOS REAIS	30
4.1	ESTUDO DE SIMULAÇÃO	30
4.2	ANÁLISE EM DADOS REAIS DE NDVI E EVI	33
4.2.1	<i>Normalized Difference Vegetation Index</i> - NDVI	34
4.2.2	<i>Enhanced Vegetation Index</i> - EVI	35
4.2.3	Análise de autocorrelação espacial	36

4.2.4	Mapas de Pernambuco - NDVI e EVI	38
4.2.5	Análise de regressão	39
5	CONCLUSÃO	45
	REFERÊNCIAS	46
	APÊNDICE A – DEDUÇÃO MATEMÁTICA DA DISTRIBUIÇÃO E PROPRIEDADES ASSOCIADAS	49
	APÊNDICE B – VETOR <i>ESCORE</i> E MATRIZ DE INFORMAÇÃO DE FISHER	56

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

A teoria de distribuições é essencial para descrever fenômenos aleatórios em diversas áreas do conhecimento. No entanto, muitas distribuições existentes apresentam limitações ao descrever certos conjuntos de dados, especialmente quando há assimetria, caudas pesadas ou comportamentos de multimodalidade. Como solução a este problema, uma série de geradores de distribuições tem sido amplamente investigada nos últimos anos. Vários desses geradores podem ser considerados casos específicos da família de geradores "transformed T-X". Nesse contexto, a família logistic-X de Tahir et al. (2016) representa uma subfamília do gerador T-X, assim como as famílias generalizadas gamma-G, conhecidas como (ZOGRAFOS; BALAKRISHNAN, 2009) e (RISTIČ; BALAKRISHNAN, 2012).

Gleaton e Lynch (2006) introduziram uma família de distribuições log-logísticas ímpares. Esta família foi derivada de uma generalização da distribuição log-logística, amplamente utilizada em dados de sobrevivência, ao incluir um parâmetro adicional de forma. Ela possibilitou criar diversas distribuições com características importantes, como funções de risco crescentes, decrescentes, unimodais, em formato de U e até bimodais. Dentro dessa abordagem, Lima et al. (2019) construíram um modelo de regressão com dispersão variável a partir da Família *Odd Log-Logistic Geometric*, para dados de censura e tempo de vida. Vagas et al. (2024) propuseram uma família generalizada de distribuições baseada na logística, conhecida como *Beta Odd Log-Logistic*, aplicando a dados reais de resistência das fibras de vidros e pacientes com câncer de bexiga. Fitzgerald (2005) explorou a aplicação dessa distribuição a chuvas extremas.

Nesta dissertação, algumas características motivarão a escolha de uma distribuição de base: (i) uma neutralidade da distribuição *baseline* tanto para impor novos parâmetros como para produzir um modelo de regressão e (ii) um suporte em $(0,2)$ para descrever o NDVI (Normalized Difference Vegetation Index) modificado a partir do EVI (Enhanced Vegetation Index). Diante disso, objetiva-se propor uma nova distribuição reparametrizada em termos da média no intervalo $(0,2)$.

1.2 CONTRIBUIÇÃO DESTA DISSERTAÇÃO

A principal contribuição desta dissertação é, em se propondo uma distribuição marginal com as características definidas no final da última seção, introduzir um novo modelo de regressão para descrever o índice de vegetação NDVI modificado, $NDVI^* = NDVI + 1$. A regressão será chamada de regressão GOLL- $\mathcal{T}$, tendo por maginal a lei GOLL- $\mathcal{T}$ reparametrizada em termo da média. Algumas propriedades da distribuição GOLL- $\mathcal{T}$ reparametrizada são derivadas, tais como: variância, curtose, assimetria e entropia de Shannon. Adicionalmente, a regressão GOLL- $\mathcal{T}$ é introduzida e algumas propriedades analíticas associadas são derivadas: versões matriciais para as funções score e matriz de informação de Fisher e o estimador de *scoring-Fisher* para os parâmetros da regressão proposta.

Do ponto de vista numérico, é elaborado um estudo de simulação Monte Carlo para quantificar a *performace* das estimativas de máxima verossimilhança em vários cenários. Os resultados das estimativas dos parâmetros β_0 , β_1 e β_2 nos diferentes cenários considerados mostram que, conforme o tamanho da amostra (n) aumenta, tanto o viés quanto o erro quadrático médio (EQM) tendem a diminuir, indicando que elas satisfazem o que se espera assintoticamente dos estimadores de máxima verossimilhança.

Adicionalmente, o modelo de regressão GOLL- $\mathcal{T}$ é aplicado para descrever o impacto da variável EVI sobre a média do $NDVI^*$, que varia de 0 (representando estado de não vegetação) a 2 (estado de vegetação densa). Resultados favorecem a proposição do método desta dissertação.

1.3 PLATAFORMA COMPUTACIONAL

1.3.1 Texto e computação

A linguagem funcional de programação escolhida para a produção de gráficos e a realização de simulações neste estudo foi o **R** (CHAMBERS, 2008). O texto foi redigido utilizando o **LaTeX**, e as referências foram gerenciadas através do **BibTeX**.

1.3.2 SATVeg - Sistema de Análise Temporal da Vegetação

O **SATVeg** é uma ferramenta web desenvolvida pela **Embrapa Agricultura Digital**, projetada para o acesso e visualização de perfis temporais dos índices vegetativos **NDVI** e **EVI**

do sensor **MODIS**. A plataforma permite analisar a vegetação de qualquer local da América do Sul, oferecendo uma interface intuitiva integrada com **Google Maps**.

Com **atualizações regulares** de dados, conforme a disponibilidade de imagens pelo **LP-DAAC/EOS-NASA**, o SATVeg oferece uma série de funcionalidades úteis como:

- Localização de áreas de interesse no mapa;
- Ferramentas para filtragem de séries temporais;
- Módulos para o carregamento de arquivos vetoriais;
- E outras ferramentas avançadas para análise.

1.3.2.1 *Dados Utilizados*

Os índices de vegetação, como o NDVI e o EVI, são ferramentas essenciais para monitorar a qualidade da vegetação a partir de dados de satélite. O NDVI destaca a diferença entre a radiação visível e infravermelha refletida pela vegetação, permitindo avaliar sua densidade e atividade fotossintética (PONZONI; SHIMABUKURO; KUPLICH, 2015). Por sua vez, o EVI pode ser entendido como uma versão aprimorada do NDVI, oferecendo maior sensibilidade em áreas com vegetação densa, minimizando os efeitos da saturação e das condições atmosféricas (ROUSE et al., 1974; HUETE et al., 2002a). Os autores Zanzarini et al. (2013) explicam que o índice NDVI varia de -1 a 1, sendo que valores próximos de 1 indicam vegetação densa e saudável, enquanto valores próximos de 0 ou negativos refletem a ausência de vegetação.

1.4 ORGANIZAÇÃO DA DISSERTAÇÃO

O desenvolvimento desta dissertação está estruturado da seguinte maneira:

- **Capítulo 2:** A distribuição GOLL- $\mathcal{T}$, seus momentos (média, variância, assimetria e curtose) e a entropia de Shannon (que fornece uma medida da incerteza associada ao modelo proposto) são abordados. Neste capítulo, também são discutidos os fundamentos matemáticos necessários para entender a distribuição.
- **Capítulo 3:** A regressão é explorada de forma abrangente, com ênfase na log-verossimilhança, a função score, a matriz de informação de Fisher e o cálculo do resíduo de

deviance. Este capítulo apresenta a abordagem inferencial completa utilizada para a análise estatística, com destaque para o método *scoring-Fisher*.

- **Capítulo 4:** A análise numérica é realizada com a simulação de Monte Carlo, a fim de quantificar a *performance* de estimação dos parâmetros do modelo propostos. Além disso, são apresentadas aplicações dos métodos a dados reais, com foco nos índices de vegetação NDVI e EVI.

- **Capítulo 5:** Considerações finais e perspectivas futuras são apresentadas.

2 DISTRIBUIÇÃO GOLL- $\mathfrak{T}$

2.1 DEFININDO A FAMÍLIA GOLL- $\mathfrak{T}$

Baseando-se no gerador de distribuição T-X proposto por Alzaatreh, Lee e Famoye (2013), Cordeiro et al. (2017b) introduziram uma nova e mais ampla classe de distribuições contínuas, denominada família log-logística ímpar generalizada (GOLL-G). Esta família possui função de distribuição acumulada (fda) definida como

$$F(x; \alpha, \gamma, \boldsymbol{\xi}) = \frac{G(x, \boldsymbol{\xi})^{\alpha\gamma}}{G(x, \boldsymbol{\xi})^{\alpha\gamma} + [1 - G(x, \boldsymbol{\xi})^\gamma]^\alpha}, \quad (2.1)$$

em que $g(x, \boldsymbol{\xi})$ é a função de densidade de probabilidade (fdp) de uma distribuição *baseline*, $G(x; \boldsymbol{\xi})$ é a fda da *baseline*, $\alpha > 0$ e $\gamma > 0$ são parâmetros de forma e $\boldsymbol{\xi} \in \mathbb{R}^m$ é o vetor de parâmetros da *baseline*. A família GOLL-G inclui como casos particulares a família log-logística ímpar (OLL), proposta por Gleaton e Lynch (2010), e a de taxa de risco proporcional reversa desenvolvida por Gupta e Gupta (2007). Adicionalmente, a fda da *baseline* G representa o caso particular quando $\alpha = \gamma = 1$. A fdp correspondente à (2.1) dada por:

$$f(x; \alpha, \gamma, \boldsymbol{\xi}) = \frac{\alpha\gamma g(x, \boldsymbol{\xi}) G(x, \boldsymbol{\xi})^{\alpha\gamma-1}}{[1 - G(x, \boldsymbol{\xi})^\gamma] \{G(x, \boldsymbol{\xi})^{\alpha\gamma} + [1 - G(x, \boldsymbol{\xi})^\gamma]^\alpha\}}. \quad (2.2)$$

Este caso é denotado como $X \sim \text{GOLL-G}(\alpha, \gamma, \boldsymbol{\xi})$.

2.2 UMA NOVA DISTRIBUIÇÃO REPARAMETRIZADA (GOLL- $\mathfrak{T}$)

Nessa seção, é apresentada uma nova distribuição reparametrizada, chamada GOLL- $\mathfrak{T}$. O resultado é obtido por usar a distribuição uniforme na família GOLL-G e construir uma reparametrização em função da média. Assim, considere:

$$g(x; \boldsymbol{\xi}) = \frac{1}{2} \mathbb{I}_{(0,2)}(x) \quad \text{e} \quad G(x; \boldsymbol{\xi}) = \frac{x}{2} \mathbb{I}_{(0,2)}(x) + \mathbb{I}_{[2,\infty)}(x), \quad (2.3)$$

fdp e fda de uma uniforme em $(0,2)$, respectivamente, em que $\mathbb{I}_A(y)$ é uma função indicadora em A.

Seja Y_0 uma variável aleatória com densidade

$$f(y; \alpha, \gamma) = \frac{\frac{\gamma \alpha}{2} \left(\frac{y}{2}\right)^{\gamma \alpha - 1} [1 - \left(\frac{y}{2}\right)^{\gamma}]^{\alpha - 1}}{\left\{ \left(\frac{y}{2}\right)^{\gamma \alpha} + [1 - \left(\frac{y}{2}\right)^{\gamma}]^{\alpha} \right\}^2} \mathbb{I}_{(0,2)}(y), \quad (2.4)$$

resultante da aplicação de (2.3) em (2.2). Para $\alpha = 1$, a Eq (2.4) se reduz a:

$$f(y; \gamma) = \frac{\gamma}{2} \left(\frac{y}{2}\right)^{\gamma - 1} = \frac{\gamma}{2^{\gamma}} y^{\gamma - 1} \mathbb{I}_{(0,2)}(y). \quad (2.5)$$

Agora, estamos em posição de propor a distribuição reparametrizada. Admita que Y_0 é uma variável aleatória tal que sua densidade é (2.5), então:

$$\begin{aligned} E(Y_0) &= \int_0^2 y \frac{\gamma}{2^{\gamma}} y^{\gamma - 1} dy = \frac{\gamma}{2^{\gamma}} \int_0^2 y^{\gamma} dy = \frac{\gamma}{2^{\gamma}} \left[\frac{y^{\gamma + 1}}{\gamma + 1} \right]_0^2 \\ &= \frac{\gamma}{2^{\gamma}} \frac{1}{\gamma + 1} 2^{\gamma + 1} = \frac{2\gamma}{\gamma + 1}. \end{aligned}$$

Daí,

$$\mu = E(Y_0) = \frac{2\gamma}{\gamma + 1} \Leftrightarrow \gamma = \frac{\mu}{2 - \mu}. \quad (2.6)$$

Substituindo (2.6) em (2.5), tem-se:

$$f(y; \mu) = \frac{\mu}{2 - \mu} 2^{-\frac{\mu}{2 - \mu}} y^{\frac{\mu}{2 - \mu} - 1} \mathbb{I}_{(0,2)}(y). \quad (2.7)$$

Para $\mu \in (0, 2)$, a fda, $F(y; \mu)$, é dada por:

$$F(y; \mu) = \begin{cases} 0, & \text{se } y \leq 0, \\ 2^{-\frac{\mu}{2 - \mu}} y^{\frac{\mu}{2 - \mu}}, & \text{se } 0 < y < 2, \\ 1, & \text{se } y \geq 2. \end{cases} \quad (2.8)$$

A partir de agora, denotamos por $Y \sim \text{GOLL-}\mathfrak{T}(\mu)$ como uma variável aleatória com fdp (2.7) e fda (2.8). A Figura 1 apresenta curvas de densidade teórica e empírica, em que a última é gerada utilizando o método de inversão a partir de (2.8). Pode-se observar que: (i) quando $\mu < 1$, a curva apresenta um decaimento exponencial; quando $\mu = 1$, a distribuição

tende a se aproximar de uma distribuição uniforme; para $\mu > 1$, a curva exibe um crescimento exponencial.

Na Figura 1, verifica-se as curvas de densidade da distribuição proposta. Para $\mu = 0.2$, observa-se um decaimento exponencial acentuado; para $\mu = 0.6$, um decaimento exponencial moderado; para $\mu = 1$, verifica-se que os pontos estão distribuídos em torno de uma reta; para $\mu = 1.4$, os pontos ao redor da curva descrevem quase uma reta; para $\mu = 1.6$, os pontos em torno da curva descrevem um crescimento exponencial; e para $\mu = 1.9$, observa-se um crescimento exponencial mais elevado dos pontos ao redor da curva.

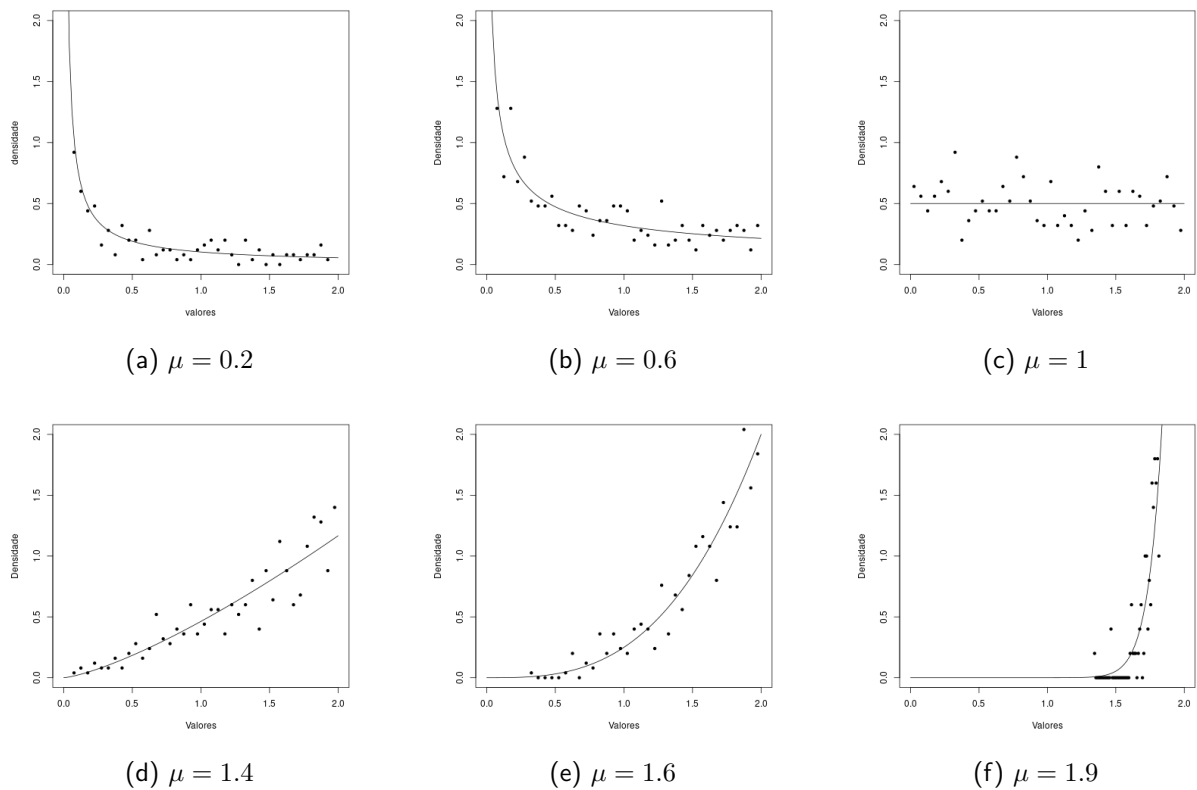


Figura 1 – Densidades da distribuição GOLL- $\mathfrak{I}$ com diferentes valores do parâmetro μ .

2.3 PROPRIEDADES DA GOLL- $\mathfrak{I}(\mu)$

Algumas propriedades da distribuição serão apresentadas a seguir.

2.3.1 Esperança e Variância

Seja $Y \sim \text{GOLL-}\mathfrak{I}(\mu)$, então sua esperança e variância são:

$$E(Y) = \int_0^2 y f(y) dy = \mu$$

e

$$\text{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{\mu(2 - \mu)^2}{4 - \mu} = \frac{\mu(4 - 4\mu + \mu^2)}{4 - \mu}.$$

Os resultados detalhados estão no Apêndice A.

A Figura 2 em particular ilustra a variância. Quando μ é pequena, a variância é próxima de zero. Quando μ aproxima de 1, a variância aumenta. Quando $\mu > 1$, a variância decai novamente.

Corolário 1. *Seja Y uma variável aleatória com distribuição GOLL- $\mathfrak{T}$. A variância máxima, $\sigma^2 = 0,36$, é obtida quando $\mu = 0,77$*

Demonstração. Precisamos mostrar que a função de variância de Y tem um único ponto crítico e que ele é o máximo dessa função. Anteriormente, foi demonstrado que

$$\text{Var}(Y) = J(\mu) \triangleq \frac{\mu(4 - 4\mu + \mu^2)}{4 - \mu} \Leftrightarrow J(\mu) = \frac{4\mu - 4\mu^2 + \mu^3}{4 - \mu}.$$

Sabe-se que, pelo teorema de Fermat, é possível obter o ponto crítico de uma função calculando sua primeira derivada e igualando a zero. Assim é preciso encontrar o valor de μ tal que $\partial J(\mu)/\partial(\mu) = 0$. Usando a regra da cadeia, tem-se que

$$\frac{\partial J(\mu)}{\partial \mu} = \frac{\partial}{\partial \mu} \left(\frac{4\mu - 4\mu^2 + \mu^3}{4 - \mu} \right) = \frac{16 - 32\mu + 16\mu^2 - 2\mu^3}{(4 - \mu)^2}. \quad (2.9)$$

Portanto, segue que

$$\frac{\partial J(\mu)}{\partial \mu} = 0 \Leftrightarrow \frac{16 - 32\mu + 16\mu^2 - 2\mu^3}{(4 - \mu)^2} = 0 \Leftrightarrow 16 - 32\mu + 16\mu^2 - 2\mu^3 = 0. \quad (2.10)$$

A equação de terceiro grau (2.10) tem raízes 2, $3 - \sqrt{5}$ e $3 + \sqrt{5}$. No entanto, como $\mu \in (0, 2)$, o valor crítico procurado é $\mu = 3 - \sqrt{5} \approx 0,77$.

Agora, é preciso mostrar que este ponto é o máximo da função de variância. Para tanto, basta mostrar que $[\partial^2 J(\mu)/\partial \mu^2]_{\mu \approx 0,77} < 0$.

Usando novamente a regra da cadeia na Equação (2.9), é possível obter que

$$\frac{\partial^2 J(\mu)}{\partial \mu^2} = \frac{2\mu^3 - 24\mu^2 + 96\mu - 96}{(4 - \mu)^3} = \frac{-32}{(4 - \mu)^3} - 2. \quad (2.11)$$

Agora substituindo na Equação (2.11) o valor de $\mu = 0,77$, tem-se que $[\partial^2 J(\mu)/\partial \mu^2]_{\mu \approx 0,77} = -1,0504 < 0$.

Logo, $\mu = 0,77$ é ponto crítico de variância e gera como valor máximo a variância $\sigma^2 \cong 0,36$.

□

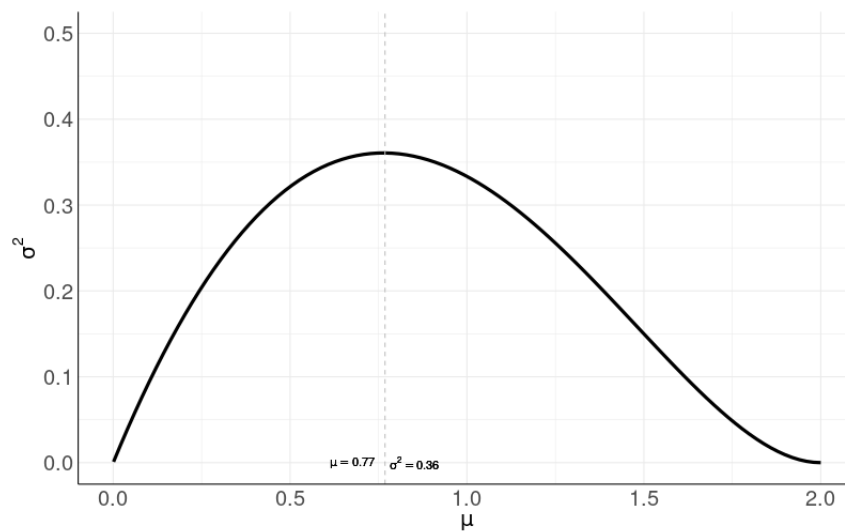


Figura 2 – Variância

2.3.2 Assimetria e Curtose

Assimetria γ_1 e Curtose γ_2 são duas medidas estatísticas fundamentais para compreender o formato de uma distribuição. A assimetria reflete o grau de simetria, já a curtose revela o comportamento das caudas, informando se os dados apresentam uma concentração maior ou menor em torno da média. A seguir, apresentamos as expressões para γ_1 e γ_2 :

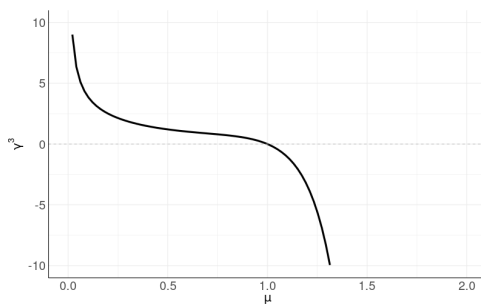
$$\gamma_1 = \frac{E[(Y - \mu)]^2}{\{E[(Y - \mu)^2]\}^{\frac{3}{2}}} = \frac{E(Y^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3} = \frac{\frac{4\mu}{3-\mu} - \frac{12\mu^2 - 12\mu^3 + 3\mu^4}{4-\mu} - \mu^3}{\frac{\mu^{\frac{3}{2}}(2-\mu)^3}{(4-\mu)^{\frac{3}{2}}}}$$

e

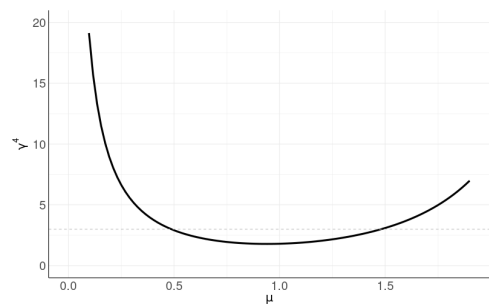
$$\begin{aligned}
\gamma_2 &= \frac{E[(Y - \mu)]^4}{(E[(Y - \mu)^2])^2} \\
&= \frac{E(Y^4) - 4\mu E(Y^3) + 6\mu^3 E(Y^2) - 4\mu^3 E(Y) + \mu^4}{\sigma^4} \\
&= \frac{\frac{16\mu}{8-3\mu} - \frac{16\mu^2}{3-\mu} + 3\mu^4 + \frac{6\mu^3(2-\mu)^2}{4-\mu}}{\frac{\mu^2(2-\mu)^4}{(4-\mu)^2}}
\end{aligned}$$

Todo cálculo de γ_1 e γ_2 é detalhado no Apêndice A.

Na Figura 3a, no ponto $\mu = 1$, a assimetria é zero ($\gamma = 0$), indicando que a distribuição é simétrica nesse valor. Conforme μ aumenta para além de 1, a assimetria decresce e se torna negativa, sugerindo uma cauda mais longa à esquerda (assimetria negativa). Para valores menores de μ , que se aproxima de zero, a assimetria cresce positivamente, indicando uma cauda mais longa à direita (assimetria positiva). Diante disso, a mudança no sinal de γ_1 em $\mu = 1$ demonstra que a distribuição varia de simétrica para assimétrica, dependendo do valor que μ assumir. Na Figura 3b, o gráfico da curtose atinge um mínimo no ponto $\mu = 1$, indicando que a distribuição é menos achatada nesse valor, logo o ponto mais baixo. Para valores de μ afastados de 1, a curtose aumenta, sugerindo que a distribuição se torna mais leptocúrtica. Sendo assim, o mínimo de curtose em $\mu = 1$ sugere que a distribuição é mais homogênea ou equilibrada nesse ponto, com menor concentração de valores em torno da média. Todavia, valores altos de curtose (γ_2) em μ próximos de 0 ou 2 indicam que a distribuição tem caudas mais extremas e é mais pontiaguda.



(a) Assimetria



(b) Curtose

Figura 3

2.4 ENTROPIA DE SHANNON PARA GOLL- $\mathfrak{T}(\mu)$

Shannon (1948) foi o pioneiro na definição de entropia e informação mútua. Nesse contexto, foi realizada uma abordagem para calcular a entropia de Shannon para uma variável aleatória

contínua Y com uma função de densidade de probabilidade $f(y)$, $H_s(Y)$. Essa entropia é uma medida da incerteza associada à distribuição de Y e definida como

$$E[-\log f(Y)] = \int_{\mathcal{Y}} -\log[f(y)]f(y) dy,$$

em que $\mathcal{Y}$ é o suporte de Y . Para o caso em questão, tem-se

$$\begin{aligned} E[-\log f(y)] &= \int_{\mathcal{Y}} -\log(f(y))f(y) dy \\ &= - \int_0^2 \left[\log \frac{\mu}{2-\mu} - \frac{\mu}{2-\mu} \log(2) + \left(\frac{\mu}{2-\mu} - 1 \right) \log(y) \right] f(y) dy \\ &= - \int_0^2 \left[\log \frac{\mu}{2-\mu} f(y) - \frac{\mu}{2-\mu} \log(2) f(y) \right. \\ &\quad \left. + \left(\frac{\mu}{2-\mu} - 1 \right) \log(y) f(y) \right] dy \\ &= -\log \frac{\mu}{2-\mu} + 1 - \frac{(2-\mu)}{\mu} + \log(2) \end{aligned} \quad (2.12)$$

O detalhamento do cálculo para entropia de Shannon, encontra-se no Apêndice A.

Na Figura 4 observa-se o comportamento da curva de entropia. Para valores de μ próximos de 0 ou 2, a entropia tende a cair, representando uma incerteza mínima. No intervalo intermediário (próximo a 1), a entropia atinge um máximo, indicando maior incerteza nesse ponto, sendo típico da entropia de Shannon.

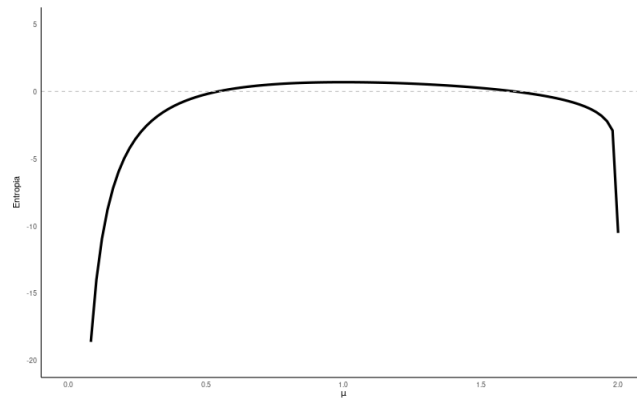


Figura 4 – Entropia de Shannon

3 MODELO DE REGRESSÃO GOLL- $\mathcal{T}$

3.1 REFERENCIAL TEÓRICO

A regressão é uma técnica estatística amplamente utilizada para modelar e analisar a relação entre uma variável dependente e uma ou mais variáveis independentes. Seu principal objetivo é entender como mudanças nas variáveis independentes podem influenciar ou prever o comportamento da variável dependente.

Inicialmente, a regressão linear foi amplamente empregada para modelar a relação entre o NDVI e variáveis de interesse, como a cobertura vegetal. No estudo de Foody (2003), foi proposta a regressão ponderada geograficamente (GWR) como uma melhoria em relação à regressão linear ordinária (OLS) para análise de dados espaciais. A pesquisa investigou a relação entre NDVI e precipitação no Norte da África e no Oriente Médio.

Posteriormente, Balaghi et al. (2008) utilizaram um modelo de regressão linear múltipla com a técnica de mínimos quadrados ordinários (OLS) para prever os rendimentos de grãos de trigo, empregando dados de NDVI, precipitação e temperatura.

No estudo de Cabrera-Bosquet et al. (2011), modelos de regressão linear foram aplicados para estabelecer relações preditivas entre o índice de vegetação NDVI e variáveis relacionadas ao crescimento da planta, como biomassa seca, área verde total, área verde sem espigas e conteúdo de nitrogênio (N) na parte aérea das plantas de trigo, considerando diferentes condições de água e nitrogênio.

De forma complementar, Propastin e Kappas (2008) compararam o uso de modelos de regressão tradicionais (OLS) e da regressão ponderada geograficamente (GWR) para modelar a relação entre o NDVI e a precipitação no Cazaquistão Central.

Mais recentemente, Li et al. (2010) utilizaram a regressão linear para modelar a relação entre o NDVI e a cobertura vegetal em diferentes tipos de vegetação, selecionando equações lineares para pastagens e arbustos, e uma equação quadrática para florestas. Similarmente, Tenreiro et al. (2021) aplicaram regressão linear para modelar a relação entre NDVI e a cobertura do dossel verde (CC) em culturas agrícolas.

Em conclusão, a regressão é uma técnica eficaz para modelar a relação entre o NDVI e variáveis vegetais, como a cobertura vegetal e o crescimento das plantas. Diferentes abordagens, incluindo a regressão ponderada geograficamente, têm sido aplicadas com sucesso para capturar variações espaciais e melhorar as previsões. A escolha do modelo depende das características

específicas dos dados e das condições do estudo, sendo essencial para alcançar resultados mais precisos e robustos na estimativa de parâmetros.

3.2 REGRESSÃO GOLL- $\mathfrak{T}$

Sejam $Y_i \sim \text{GOLL-}\mathfrak{T}$ para $i = 1, \dots, n$, em que n é o número de observações, e $\beta = [\beta_1, \dots, \beta_p]^\top$. Então, Y_i possui densidade:

$$f(y_i; \beta) = \frac{\mu_i(\beta)}{2 - \mu_i(\beta)} 2^{-\frac{\mu_i(\beta)}{2 - \mu_i(\beta)}} y_i^{\frac{\mu_i(\beta)}{2 - \mu_i(\beta)} - 1} \mathbb{I}_{(0,2)}(y_i), \quad (3.1)$$

em que

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} \Leftrightarrow \mu_i = g^{-1} \left(\beta_0 + \sum_{k=1}^p \beta_k x_{ik} \right).$$

Aqui, $g(\cdot)$ é a função de ligação que estabelece a relação entre o componente aleatório e o componente sistemático, ou seja, ela conecta a média ao preditor linear. Nesta dissertação, adaptando a função $\text{logit}(\cdot)$, utiliza-se a seguinte ligação:

$$g(\mu) = \log \left(\frac{\frac{\mu}{2}}{1 - \frac{\mu}{2}} \right).$$

Considerando $g(\mu) = \eta$, temos $\mu = \frac{2e^\eta}{1+e^\eta}$. Neste caso, tem-se que $\text{Var}(Y_i|x_i) = \frac{\mu_i(2-\mu_i)^2}{4-\mu_i}$, em que a variância depende exclusivamente da média μ e $\lim_{\mu_i \rightarrow \{0,2\}} \text{Var}(Y_i|x_i) = 0$

3.3 ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA

3.3.1 Log-Verossimilhança

Considere uma amostra aleatória ¹ $Y_1, \dots, Y_n$ tal que $Y_i \sim \text{GOLL-}\mathfrak{T}(\beta)$. Da Expressão (3.1), o logaritmo da função de verossimilhança (função de log-verossimilhança) é dada por

$$\begin{aligned} \ell(\beta) &= \log \prod_{i=1}^n f(y_i; \beta) = \prod_{i=1}^n \frac{\mu_i(\beta)}{2 - \mu_i(\beta)} 2^{-\frac{\mu_i(\beta)}{2 - \mu_i(\beta)}} \cdot y_i^{\frac{\mu_i(\beta)}{2 - \mu_i(\beta)} - 1} \\ &= \sum_{i=1}^n \log \mu_i(\beta) - \sum_{i=1}^n \log[2 - \mu_i(\beta)] - (\log 2) \sum_{i=1}^n \frac{\mu_i(\beta)}{2 - \mu_i(\beta)} \\ &\quad + \sum_{i=1}^n \frac{\mu_i(\beta)}{2 - \mu_i(\beta)} \log y_i - \sum_{i=1}^n \log y_i, \end{aligned}$$

¹ A partir de agora entenderemos "amostra aleatória" como um conjunto de variáveis aleatórias independentes e identicamente distribuídas.

em que, $y_1, \dots, y_n$ representa uma amostra observada. Assim, o estimador de máxima verossimilhança é dado por :

$$\hat{\beta} = \arg \max_{\beta \in \mathbb{R}^{p+1}} [\ell(\beta)].$$

As funções *score* obtidas a partir da log-verossimilhança são derivadas a seguir.

3.3.2 Função *Score*

A função *score* para i -ésima observação é obtida pela diferenciação da i -ésima parcela da log-verossimilhança:

$$U_i(\beta) = \frac{\partial \ell_i(\beta)}{\partial \beta} = \frac{\partial \ell_i(\beta)}{\partial \mu_i(\beta)} \frac{\partial \mu_i(\beta)}{\partial \eta_i(\beta)} \frac{\partial \eta_i(\beta)}{\partial \beta}, \quad \text{para } i = 1, \dots, n,$$

ou, equivalentemente,

$$\begin{aligned} U_i(\beta) &= \mathbf{x}_i \frac{1}{g'(\mu_i)} \left\{ \frac{1}{\mu_i(\beta)} + \frac{1}{2 - \mu_i(\beta)} + \frac{2}{[2 - \mu_i(\beta)]^2} (\log y_i - \log 2) \right\} \\ &= \mathbf{x}_i \frac{1}{g'(\mu_i)} \left\{ \frac{2}{[2 - \mu_i(\beta)]^2} \left[\log y_i - \left(\frac{2 - \mu_i}{\mu_i} - \log 2 \right) \right] \right\}, \end{aligned}$$

em que $\frac{1}{g'(\mu_i)} = \frac{\mu_i(2 - \mu_i)}{2}$. Considerando

$$\delta_i(\beta) = \frac{2}{[2 - \mu_i(\beta)]^2} \quad \text{e} \quad p_i(\beta) = -\frac{2 - \mu_i}{\mu_i} + \log 2,$$

a *score* $U_i(\beta)$ pode ser rescrita como:

$$U(\beta) = \sum_{i=1}^n U_i(\beta) \quad \text{tal que} \quad U_i(\beta) = \mathbf{x}_i \frac{1}{g'(\mu_i)} \delta_i(\beta) [\log y_i - p_i(\beta)].$$

A *score* em forma matricial sendo expressa por:

$$U(\beta) = \sum_{i=1}^n \mathbf{x}_i \frac{1}{g'(\eta_i)} \delta_i(\beta) [\log y_i - p_i(\beta)] = \mathbf{X}^\top \mathbf{W} \Delta [\log \mathbf{y} - \mathbf{p}(\beta)],$$

em que $\mathbf{X} = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid \dots \mid \mathbf{x}_n]^\top$ é a matriz modelo $n \times (p + 1)$ tal que $\mathbf{x}_i$ representa o vetor de covariáveis para a i -ésima observação, $\mathbf{W} = \text{diag}\left(\frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)}\right)$, tal que $\text{diag}(\cdot)$

representa uma matriz diagonal, $\Delta = \text{diag}(\delta_1(\beta), \dots, \delta_n(\beta))$, $\log \mathbf{y} \equiv [\log y_1, \dots, \log y_n]^\top$ e $\mathbf{p}(\beta) = [p_1(\beta), \dots, p_n(\beta)]^\top$.

Assim, de $\mathbf{U}(\beta) = \mathbf{0}$, segue-se que:

$$\mathbf{X}^\top \mathbf{W} \Delta \log \mathbf{y} = \mathbf{X}^\top \mathbf{W} \Delta \mathbf{p}(\beta).$$

3.3.3 Matriz Informação de Fisher

A matriz informação de Fisher (MIF) é uma ferramenta crucial na Estatística, particularmente em modelos lineares generalizados (MLGs) e regressão não normais. A MIF, $\mathbf{K}(\beta)$, é dada por (CASELLA; BERGER, 2002):

$$\mathbf{K}(\beta) = E \left[\frac{\partial \ell(\beta)}{\partial \beta} \frac{\partial \ell(\beta)}{\partial \beta^\top} \right] = E \left[-\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^\top} \right]. \quad (3.2)$$

A MIF desempenha um papel essencial na inferência estatística, fornecendo uma medida da quantidade de informação que uma amostra contém sobre um parâmetro desconhecido. Para que a Equação (3.2) seja válida, algumas condições de regularidade devem ser atendidas: (i) a continuidade das derivadas da função de verossimilhança, (ii) a integrabilidade da função de verossimilhança, (iii) o fato de que a derivada esperada da log-verossimilhança em relação aos parâmetros seja zero. Essas condições garantem a aplicação correta da matriz para o cálculo dos erros padrão dos estimadores de máxima verossimilhança.

Inicialmente, considere a derivação da matriz Hessiana:

$$\begin{aligned} \frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^\top} &= \frac{\partial}{\partial \beta^\top} \mathbf{U}(\beta) = \frac{\partial}{\partial \beta^\top} (\mathbf{X}^\top \mathbf{W} \Delta [\log \mathbf{y} - \mathbf{p}(\beta)]) \\ &= -\mathbf{X}^\top \mathbf{W} \Delta \frac{\partial \mathbf{p}}{\partial \beta^\top} + \mathbf{X}^\top \left[\frac{\partial(\mathbf{W} \Delta)[\log \mathbf{y} - \mathbf{p}(\beta)]}{\partial \beta_0} \mid \dots \mid \frac{\partial(\mathbf{W} \Delta)[\log \mathbf{y} - \mathbf{p}(\beta)]}{\partial \beta_p} \right], \end{aligned}$$

em que

$$\begin{aligned} \frac{\partial \mathbf{p}_i(\beta)}{\partial \beta^\top} &= -\frac{1}{2} \frac{\partial}{\partial \beta^\top} \left\{ \frac{[2 - \mu_i(\beta)]^2}{\mu_i(\beta)} \right\} - \frac{1}{2} \frac{\partial [2 - \mu_i(\beta)]}{\partial \beta^\top} \\ &= -\frac{1}{2} \left\{ \frac{-2 \frac{\partial \mu_i(\beta)}{\partial \beta^\top} [2 - \mu_i(\beta)] - \frac{\partial \mu_i(\beta)}{\partial \beta^\top} [2 - \mu_i(\beta)]^2}{\mu_i(\beta)^2} \right\} + \frac{1}{2} \frac{\partial \mu_i(\beta)}{\partial \beta^\top} \end{aligned}$$

$$= \frac{1}{2} \frac{\partial \mu_i(\beta)}{\partial \beta^\top} + \frac{1}{2} \frac{[2 - \mu_i(\beta)]}{\mu_i^2(\beta)} \frac{\partial \mu_i(\beta)}{\partial \beta^\top} [1 + 2 - \mu_i(\beta)],$$

$$\begin{aligned} \frac{\partial(\mathbf{W}\Delta)}{\partial \beta_k} &= \left\{ \text{dig} \left[\left(\frac{\partial \frac{1}{g'(\mu_1)}}{\partial \beta_k} \right) \delta_1 + \frac{1}{g'(\mu_1)} \frac{\partial \delta_1}{\partial \beta_k}, \dots \right. \right. \\ &\quad \left. \left. , \left(\frac{\partial \frac{1}{g'(\mu_n)}}{\partial \beta_k} \right) \delta_n + \frac{1}{g'(\mu_n)} \frac{\partial \delta_n}{\partial \beta_k} \right] \right\}, \\ \frac{\partial \frac{1}{g'(\mu_i)}}{\partial \beta_k} &= \frac{\partial \frac{1}{g'(\mu_i)}}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_k} = -\frac{g''(\mu_i)}{[g'(\mu_i)]^2} \mathbf{x}_{ik} \end{aligned}$$

e

$$\frac{\partial \delta_i}{\partial \beta_k} = \frac{4[2 - \mu_i(\beta)] \frac{\partial \mu_i(\beta)}{\partial \beta_k}}{[2 - \mu_i(\beta)]^4} = \frac{4[2 - \mu_i(\beta)]}{[2 - \mu_i(\beta)]^4} \frac{\partial \mu_i(\beta)}{\partial \mu_i(\beta)} \frac{\partial \mu_i}{\partial \beta_k} = \frac{4}{[2 - \mu_i(\beta)]^3} \frac{1}{g'(\mu_i)} x_{ik}.$$

Assim, tem-se que:

$$\begin{aligned} \mathbf{K}(\beta) &= -E \left[-\mathbf{X}^\top \mathbf{W} \Delta \frac{\partial \mathbf{p}}{\partial \beta^\top} + \mathbf{X}^\top \frac{\partial(\mathbf{W}\Delta)}{\partial \beta^\top} (\log \mathbf{y} - \mathbf{p}(\beta)) \right] \\ &= \mathbf{X}^\top \mathbf{W} \Delta \frac{\partial \mathbf{p}}{\partial \beta^\top} - \mathbf{X}^\top \frac{\partial(\mathbf{W}\Delta)}{\partial \beta^\top} E(\log \mathbf{y} - \mathbf{p}(\beta)) \\ &= \mathbf{X}^\top \mathbf{W} \Delta \mathbf{V}, \end{aligned}$$

em que $\mathbf{V} = \frac{\partial \mathbf{p}(\beta)}{\partial \beta^\top}$ é uma matriz $n \times (p+1)$, dada por:

$$\mathbf{V} = \begin{bmatrix} \frac{2}{\mu_1^2} \frac{2-\mu_1}{\mu_1} \mathbf{x}_1^\top \\ \frac{2}{\mu_2^2} \frac{2-\mu_2}{\mu_2} \mathbf{x}_2^\top \\ \vdots \\ \frac{2}{\mu_n^2} \frac{2-\mu_n}{\mu_n} \mathbf{x}_n^\top \end{bmatrix}.$$

3.3.4 Método *Score*-Fisher

O EMV do vetor β não possui uma expressão analítica fechada. Assim, é necessário utilizar um método numérico, como o de Newton-Raphson ou o escore de Fisher, para estimar β . De

acordo com Paula (2004), para estimar os parâmetros, costuma-se utilizar o método iterativo de Newton-Raphson. É possível obtê-lo expandindo a função escore $U(\beta)$ em torno de $\beta^{(0)}$,

$$U(\beta) \cong U(\beta^{(0)}) + U'(\beta^{(0)})(\beta - (\beta^{(0)})),$$

em que U' expressa a primeira derivada de $U(\beta)$ em relação a β . A partir de uma manipulação analítica simples, a seguinte equação iterativa é obtida:

$$\beta^{(m+1)} = \beta^{(m)} + \{-U'(\beta^{(m)})\}^{-1} U(\beta^{(m)}),$$

para $m = 0, 1, \dots$. Como a matriz $-U'(\beta^{(m)})$ pode não ser positiva definida, o método de *scoring* de Fisher (MEF), sugere substituir essa matriz pelo valor esperado correspondente. Esse procedimento resulta no seguinte processo iterativo:

$$\beta^{(m+1)} = \beta^{(m)} + K^{-1}(\beta^{(m)}) U(\beta^{(m)}).$$

O procedimento de estimação de β com o método do *scoring* de Fisher fica dado por:

$$\beta^{(m+1)} = \beta^{(m)} + \left[X^T W^{(m)} \Delta^{(m)} V^{(m)} \right]^{-1} X^T W^{(m)} \Delta^{(m)} [\log y - p(\beta^{(m)})]. \quad (3.3)$$

3.3.5 Resíduo *Deviance*

De acordo com Paula (2004), a qualidade do ajuste de um MLG pode ser avaliada por meio do resíduo *deviance*. Essa quantidade mede a diferença entre o logaritmo da verossimilhança do modelo saturado e do modelo em estudo, utilizando a estimativa de máxima verossimilhança dos coeficientes. Valores menores para o resíduo *deviance* indicam que o modelo simplificado apresenta um ajuste comparável ao modelo saturado. Matematicamente, o resíduo *deviance* é definido como:

$$D^*(y; \hat{\mu}) = \phi D(y; \hat{\mu}) = 2\{L(y; y) - L(\hat{\mu}; y)\}.$$

Além disso, uma formulação alternativa para $D(y; \hat{\mu})$ é expressa por:

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \{y_i(\hat{\theta}_i^0 - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\hat{\theta}_i^0))\}.$$

4 ESTUDO DE SIMULAÇÃO E APLICAÇÃO EM DADOS REAIS

Neste capítulo, realizam-se um estudo de simulação e aplicação a dados reais. No caso do estudo de simulação, uma análise numérica é realizada com intuito de analisar o desempenho das estimativas de máxima verossimilhança (MV) dos parâmetros do modelo de regressão GOLL- $\mathfrak{T}$. Através de simulação Monte Carlo, diferentes cenários são considerados para os parâmetros. A aplicação do modelo proposto aos índices de vegetação NDVI* e EVI será abordada em etapas posteriores, com o intuito de validar o modelo em dados reais de monitoramento ambiental.

4.1 ESTUDO DE SIMULAÇÃO

Nesta seção, será avaliado o desempenho das estimativas por meio de simulação Monte Carlo. Os valores das covariáveis foram obtidas da distribuição uniforme $X \sim U[-1, 1]$ para tamanhos de amostra iguais a 10, 20, 40 e 100, e, para cada tamanho de amostra, foram utilizadas 5.000 réplicas. Todas as simulações foram executadas com linguagem de programação funcional R. Para cada replicação, gerou-se a variável dependente $Y = [Y_1, \dots, Y_n]^T$ com $Y_i \sim \text{GOLL-}\mathfrak{T}(\beta_0, \beta_1)$ tal que $E(Y_i|X = x_i) = \frac{2e^{\eta_i}}{1+e^{\eta_i}}$, e $\eta_i = \beta_0 + \beta_1 x_i$.

Tabela 1 – Resultados da simulação com diferentes tamanhos da amostra quando os parâmetros $\beta_0 = 1$ e $\beta_1 = -2$, e $\bar{\beta}$ é a média das estimativas e $\hat{b}_{\bar{\beta}}$ é o viés.

β	n	$\bar{\beta}$	$\hat{b}_{\bar{\beta}}$	EQM
$\beta_0 = 1$	10	1.1146	0.1146	2.9039×10^{-5}
	20	1.0525	0.0525	1.1787×10^{-5}
	40	1.0234	0.0234	5.3621×10^{-6}
	100	1.0098	0.0098	2.1391×10^{-6}
$\beta_1 = -2$	10	-2.0080	-0.0080	9.1076×10^{-5}
	20	-1.9953	0.0046	3.5116×10^{-5}
	40	-2.0851	-0.0085	1.5951×10^{-5}
	100	-2.0035	-0.0035	6.1231×10^{-6}

A Tabela 1 apresenta os resultados para os diferentes tamanhos da amostra quando $\beta_0 = 1$ e $\beta_1 = -2$. Ao observar os resultados para o parâmetro β_0 , percebe-se que o viés diminui com o aumento do tamanho da amostra, por exemplo de 0.1146 quando $n = 10$ diminui para 0.0098 quando $n = 100$. O Erro Quadrático Médio (EQM) também diminui, de 2.9039×10^{-5} para 2.1391×10^{-6} quando o tamanho da amostra vai de 10 para 100. Isto indica que os valores das estimativas médias se aproximam do valor verdadeiro do parâmetro com o aumento da

amostra. Para o parâmetro β_1 , o viés de -0.0080 quando $n = 10$ diminui para -0.0035 quando $n = 100$, evidenciando um comportamento similar ao observado para β_0 . O EQM para β_1 também apresenta uma diminuição de 9.1076×10^{-5} para 6.1231×10^{-6} conforme o tamanho da amostra aumenta. Esses resultados indicam que tanto para β_0 quanto para β_1 , o viés e o EQM diminuem com o aumento do tamanho da amostra.

Tabela 2 – Resultados da simulação com diferentes tamanhos da amostra quando os parâmetros $\beta_0 = 1$ e $\beta_1 = 2$, e $\bar{\beta}$ é a média das estimativas e $\hat{b}_{\bar{\beta}}$ é o viés.

β	n	$\bar{\beta}$	$\hat{b}_{\bar{\beta}}$	EQM
$\beta_0 = 1$	10	1.1146	0.1146	2.9039×10^{-5}
	20	1.0525	0.0525	1.1787×10^{-5}
	40	1.0234	0.0234	5.3621×10^{-6}
	100	1.0098	0.0098	2.1391×10^{-6}
$\beta_1 = 2$	10	1.9919	-0.0080	9.1076×10^{-5}
	20	2.0046	0.0046	3.5116×10^{-5}
	40	1.9911	-0.0081	1.5951×10^{-5}
	100	1.9964	-0.0035	6.1231×10^{-6}

A Tabela 2 apresenta os resultados das simulações para diferentes tamanhos amostrais, considerando os parâmetros $\beta_0 = 1$ e $\beta_1 = 2$. Para o parâmetro β_0 , observa-se que o viés diminuem com o aumento do tamanho da amostra, variando de 0.1146 para $n = 10$ até 0.0098 para $n = 100$. O EQM também diminuem, passando de 2.9039×10^{-5} para 2.1391×10^{-6} , evidenciando o ganho de precisão com o aumento amostral. Em relação ao parâmetro β_1 , o viés diminuem de -0.0080 para $n = 10$ até -0.0035 para $n = 100$, comportamento semelhante ao β_0 . O EQM também apresenta diminuição, variando de 9.1076×10^{-5} para 6.1231×10^{-6} . Tais Resultados indicam que, para ambos os parâmetros, o viés e o EQM diminuem com o aumento da amostra, refletindo maior acurácia e consistência dos estimadores.

A Tabela 3 apresenta os resultados das simulações para distintos tamanhos amostrais, considerando os parâmetros $\beta_0 = 1$, $\beta_1 = 2$ e $\beta_2 = -1$. Para o parâmetro β_0 , observa-se que o viés diminui com o aumento do tamanho da amostra, variando de 0.1729 para $n = 10$ até 0.0167 para $n = 100$. O EQM também diminui, passando de 3.9294×10^{-4} para 2.2001×10^{-6} , indicando aumento na precisão das estimativas.

No caso de β_1 , o viés diminui de 0.0145 para -0.0018, enquanto o EQM passa de 1.7369×10^{-4} para 6.2241×10^{-6} . Para o parâmetro β_2 , observa-se a diminuição do viés de -0.0058 quando $n = 10$ para -8.4426×10^{-6} quando $n = 100$, além da diminuição do EQM de 1.001×10^{-4} para 6.4501×10^{-6} . Dos resultados obtidos dos parâmetros β_0 , β_1 e β_2 ,

Tabela 3 – Resultados da simulação com diferentes tamanhos da amostra quando os parâmetros $\beta_0 = 1$, $\beta_1 = -2$ e $\beta_2 = -1$, e $\bar{\beta}$ é a média das estimativas e $\hat{b}_{\bar{\beta}}$ é o viés.

β	n	$\bar{\beta}$	$\hat{b}_{\bar{\beta}}$	EQM
$\beta_0 = 1$	10	1.1729	0.1729	3.9294×10^{-5}
	20	1.0812	0.0812	1.3658×10^{-5}
	40	1.0391	0.0391	5.5741×10^{-6}
	100	1.0167	0.0167	2.2001×10^{-6}
$\beta_1 = -2$	10	-1.9854	0.0145	1.0001×10^{-4}
	20	-2.0118	-0.0118	3.9634×10^{-5}
	40	-2.0024	-0.0024	1.7369×10^{-5}
	100	-2.0018	-0.0018	6.2241×10^{-6}
$\beta_2 = -1$	10	-1.0058	-0.0058	1.0001×10^{-4}
	20	-0.9997	0.0002	3.9553×10^{-5}
	40	-0.9975	0.0024	1.7553×10^{-5}
	100	-1.0001	-8.4426×10^{-6}	6.4501×10^{-6}

é possível observar que a média das estimativas, o viés e o EQM diminuem à medida que o tamanho da amostra aumenta, convergindo para os respectivos valores verdadeiros.

Tabela 4 – Resultados da simulação com diferentes tamanhos da amostra quando os parâmetros $\beta_0 = 1$, $\beta_1 = 2$ e $\beta_2 = 1$ e $\bar{\beta}$ é a média das estimativas e $\hat{b}_{\bar{\beta}}$ é o viés.

β	n	$\bar{\beta}$	$\hat{b}_{\bar{\beta}}$	EQM
$\beta_0 = 1$	10	1.1729	0.1729	3.9294×10^{-5}
	20	1.0812	0.0812	1.3658×10^{-5}
	40	1.0391	0.0391	5.5741×10^{-6}
	100	1.0167	0.0167	2.2001×10^{-6}
$\beta_1 = 2$	10	2.0145	0.0145	1.0001×10^{-4}
	20	1.9881	-0.0118	3.9635×10^{-5}
	40	1.9975	-0.0024	1.7369×10^{-5}
	100	1.9981	-0.0018	6.224×10^{-6}
$\beta_2 = 1$	10	0.9941	-0.0058	1.0001×10^{-4}
	20	1.0002	0.0002	3.9554×10^{-5}
	40	1.0024	0.0024	1.7553×10^{-5}
	100	0.9999	-8.6252×10^{-6}	6.4501×10^{-6}

A Tabela 4 apresenta os resultados das simulações considerando diferentes tamanhos amostrais para os parâmetros $\beta_0 = 1$, $\beta_1 = 2$ e $\beta_2 = 1$. A média das estimativas ($\bar{\beta}$) obtidas em cada cenário indica o comportamento médio do estimador ao longo das simulações. Essa média funciona como uma aproximação empírica do valor esperado do estimador, sendo fundamental para a avaliação tanto do viés quanto da precisão.

Para o parâmetro β_0 , observa-se que a média das estimativas $\bar{\beta}_0$ se aproxima progressivamente do valor verdadeiro à medida que o tamanho da amostra aumenta, passando de 1.1729 quando $n = 10$ para 1.0167 quando $n = 100$. Essa aproximação é acompanhada por uma redução do viés de 0.1729 para 0.0167. E o EQM que passa de 3.9294×10^{-5} para 2.2001×10^{-6} , evidenciando o aumento da precisão e a consistência do estimador.

Para o parâmetro β_1 , a média $\bar{\beta}_1$ também se estabiliza em torno do valor verdadeiro (2), com o viés diminuindo de 0.0145 para -0.0018 conforme n passa de 10 para 100. O EQM apresenta comportamento semelhante, diminuindo de 1.0001×10^{-4} para 6.2241×10^{-6} , o que indica uma melhoria significativa na precisão do estimador. A média aqui exerce papel central como estimativa do valor esperado para cada tamanho amostral.

Já para o parâmetro β_2 , observa-se comportamento semelhante: $\bar{\beta}_2$ evolui de 0.9941 para 0.9999, com o viés diminuindo de -0.0058 para aproximadamente zero $-8,6252 \times 10^{-6}$ e o EQM de 1.0001×10^{-4} para 6.4501×10^{-6} . A média das estimativas também se mostra eficiente como indicativo de convergência para o valor verdadeiro do parâmetro.

Portanto, os resultados demonstram que, para todos os parâmetros analisados, a média das estimativas se aproxima dos verdadeiros valores conforme o tamanho da amostra aumenta. Essa convergência é acompanhada pela redução do viés e do EQM, indicando que os estimadores utilizados são consistentes e tornam-se mais precisos à medida que o tamanho da amostra cresce.

4.2 ANÁLISE EM DADOS REAIS DE NDVI E EVI

Índices como o NDVI e o EVI são amplamente utilizadas para monitorar a qualidade da vegetação a partir de dados coletados via satélite. O NDVI apresenta limitação em áreas com vegetação muito densa, como florestas tropicais, onde tende a saturar. Isso significa que o índice atinge valores próximos do máximo e perde a capacidade de diferenciar variações na estrutura da vegetação. Por outro lado, o EVI, desenvolvido para contornar essa limitação, mantém sua sensibilidade mesmo em regiões com alta biomassa, permitindo detectar variações na cobertura vegetal superior, como as copas das árvores (HUETE et al., 2002a). O uso dessas imagens possibilita avaliações mais precisas, contribuindo para decisões informadas em áreas como o manejo agrícola e a conservação ambiental, explorado em detalhes por Fung e Siu (2000), Damian et al. (2019).

4.2.1 *Normalized Difference Vegetation Index - NDVI*

Dentre os índices desenvolvidos para o monitoramento da vegetação, o NDVI é obtido como:

$$NDVI = \frac{IR - R}{IR + R},$$

em que, IR corresponde à radiação na faixa infravermelha do espectro eletromagnético e R representa a radiação na faixa vermelha (ROUSE et al., 1974; LEE, 2005). Os valores do NDVI variam entre -1 e 1. Valores próximos de -1 estão associados a áreas com pouca ou nenhuma vegetação, como solos descobertos, rochas ou água. Por outro lado, valores positivos indicam presença de vegetação, com valores mais altos indicando maior densidade de vegetação saudável (CORDEIRO et al., 2017a).

Os dados de NDVI mostram que, em áreas com vegetação densa, como florestas, ocorre uma maior reflexão de luz, especialmente em uma faixa de luz invisível, característica de vegetação saudável. Em contraste, superfícies sem vegetação, como solos expostos ou construções, tendem a absorver a maior parte dessa radiação. Já em regiões com vegetação moderada, o comportamento é intermediário, com uma reflexão parcial da luz, dependendo da densidade e do tipo de vegetação presente (ACCIOLY et al., 2002; WELCH et al., 2005; KARIYEVA; LEEUWEN, 2011). A faixa de luz invisível (ou infravermelho) é uma parte do espectro eletromagnético que o olho humano não consegue perceber, mas que desempenha um papel crucial na análise da vegetação, pois plantas saudáveis refletem mais essa luz (HUETE et al., 2002a).

O NDVI é utilizado em diversas aplicações, como a detecção de áreas com vegetação, estimativas de biomassa, análise do crescimento e desempenho das plantas, além de avaliação da cobertura vegetal. Este índice tem se mostrado eficaz na estimativa da biomassa acima do solo, já que sua correlação com a densidade e o desenvolvimento das plantas permite uma avaliação precisa e eficiente. (ZHU; LIU, 2015). Por exemplo, em um canavial, o NDVI foi utilizado para identificar áreas cultivadas e avaliar sua condição, fornecendo informações sobre a qualidade da vegetação, a cobertura da plantação e o crescimento das plantas (RAHMAN; ISLAM; RAHMAN, 2004).

4.2.2 *Enhanced Vegetation Index - EVI*

Proposto por Huete et al. (2002b), o EVI foi desenvolvido para otimizar a detecção da vegetação, proporcionando maior sensibilidade em áreas de alta biomassa e minimizando as influências atmosféricas e o efeito do solo. Essa abordagem é possível devido à incorporação de um termo de resistência ao aerossol¹, que utiliza a banda azul para corrigir as interferências na banda vermelha. A equação do EVI é expressa por:

$$EVI = G \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_1 \rho_{red} - C_2 \rho_{blue} + L},$$

em que, ρ_{NIR} , ρ_{red} , ρ_{blue} representam as refletâncias corrigidas atmosféricamente das bandas espectrais, C_1 , C_2 são coeficientes de correção de aerossóis, L é um fator de ajuste para minimizar efeitos do solo, e G é um fator de ganho. Os valores comumente utilizados são $C_1 = 6$, $C_2 = 7.5$, $L = 10000$, $G = 2.5$ (HUETE; JUSTICE; LIU, 1994; HUETE et al., 1997).

O Índice de Vegetação Aprimorado, EVI, é amplamente utilizado para avaliar a qualidade das plantações, monitorar mudanças no uso do solo e estimar a biomassa em ecossistemas naturais e agrícolas. Sua capacidade de minimizar os efeitos atmosféricos e de solo torna-o particularmente útil para análises em regiões com alta variação de condições climáticas e geográficas. Em comparação com o NDVI, o EVI é mais sensível em áreas de alta biomassa, oferecendo uma alternativa mais precisa para o monitoramento da vegetação em grandes escalas.

O EVI tem sido relevante para a agricultura de precisão, sendo utilizado para monitorar a condição das culturas agrícolas. Segundo Kurbanov e Zakharova (2020), o EVI desempenha um papel importante na avaliação da qualidade das plantações, como a soja e o trigo, durante as principais fases de crescimento. A análise desse índice é fundamental para detectar variações no crescimento das plantas, fornecendo informações sobre o estado das plantações e facilitando a tomada de decisões agrônômicas. Em um estudo realizado por Silva et al. (2020), o EVI foi eficaz na classificação da cobertura vegetal em Campo Belo do Sul, estado de Santa Catarina. Os resultados mostraram que o EVI, combinado com outros índices como o NDVI e o SAVI (ver detalhes em SILVA et al., 2020), teve bom desempenho na detecção das classes de vegetação. De acordo com Glenn et al. (2011), o EVI também tem sido utilizado para

¹ Aerossóis são partículas microscópicas suspensas na atmosfera, como poeira, fumaça e poluição, que podem interferir na medição de parâmetros ambientais, como a vegetação. Essas partículas podem alterar a quantidade de luz que chega aos sensores, prejudicando a precisão das imagens de satélite.

estimar a biomassa em ecossistemas naturais e agrícolas. O uso de índices como o EVI contribui para estimativas precisas da biomassa e da quantidade de água que as plantas precisam, possibilitando uma gestão eficiente da irrigação e do uso da água, tanto em sistemas agrícolas quanto em ecossistemas naturais.

Assim, o EVI se destaca como um índice essencial para o monitoramento da vegetação, contribuindo para a gestão eficiente de recursos naturais e a otimização da produção agrícola.

A obtenção dos dados para este estudo foi via requisição por API do SATVEG - Embrapa <<https://www.agroapi.cnptia.embrapa.br/store/apis/info?name=SATVeg&version=v2&provider=agroapi>> na data de 03 de dezembro de 2024. Os dados obtidos são referentes aos 184 municípios de Pernambuco mais o arquipélago de Fernando de Noronha. Foram obtidos para o NDVI 1082 observações no período de 18.02.2000 até 23.10.2024. A obtenção média do valor do NDVI em intervalos de dias para cada município é aproximadamente 8.3 dias em cada captura e 8 dias se observando a mediana. Para o EVI, o período de obtenção dos dados foram os mesmos do NDVI, 1082 observações no período de 18.02.2000 até 23.10.2024. Obtenção média do valor do EVI é similar ao do NDVI, sendo 8.3 dias em cada captura para cada município. O NDVI foi analisado utilizando valores acima do quantil 0.99, ou seja, com foco nas regiões de vegetação densa, como florestas tropicais e matas nativas, onde a cobertura do solo é quase total. Para o EVI, foi adotada a mediana. O centroide foi o ponto exato que foi extraído o NDVI e EVI em cada município mais o arquipélago de Fernando de Noronha, esses valores são obtidos de forma contínua em todo espaço geográfico, logo, a importância de saber a localização exata da captura dos dados.

A informação do NDVI e EVI é mapeada para toda região geográfica, em nosso estudo, utilizando uma localização pontual para obter o valor do NDVI e EVI para cada município. Essa localização pontual utilizada foi o centroide de cada município. Os centroides podem ser observados na Figura 5.

4.2.3 Análise de autocorrelação espacial

A Figura 6a representa a análise da autocorrelação espacial do Índice de Moran Global para os municípios de Pernambuco, incluindo o arquipélago de Fernando de Noronha. O objetivo é verificar a distribuição espacial NDVI, e identificar se seus valores estão agrupados ou dispersos geograficamente.

A estatística do Índice de Moran obtida foi: $I = 0.0926$, sendo o $p - valor = 0.0348$, que

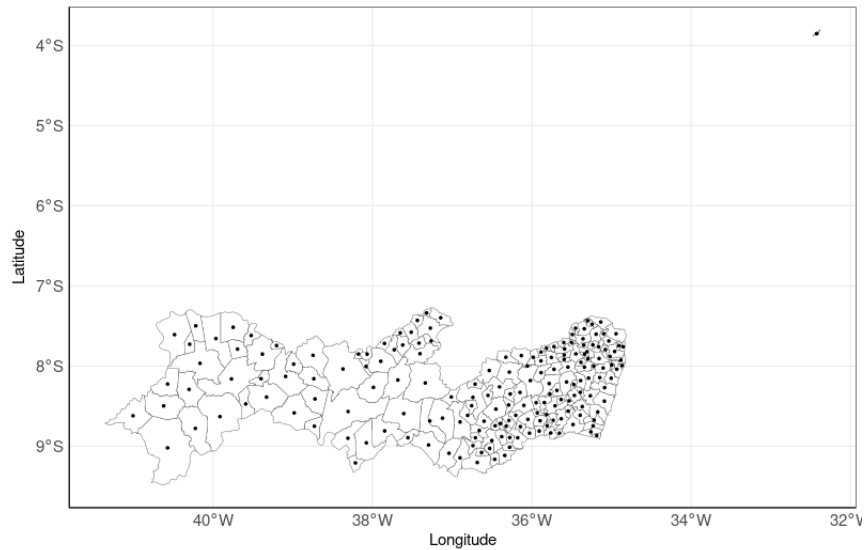


Figura 5 – Mapa de Pernambuco com localização pontual (centroide) de cada município mais o arquipélago de Fernando de Noronha.

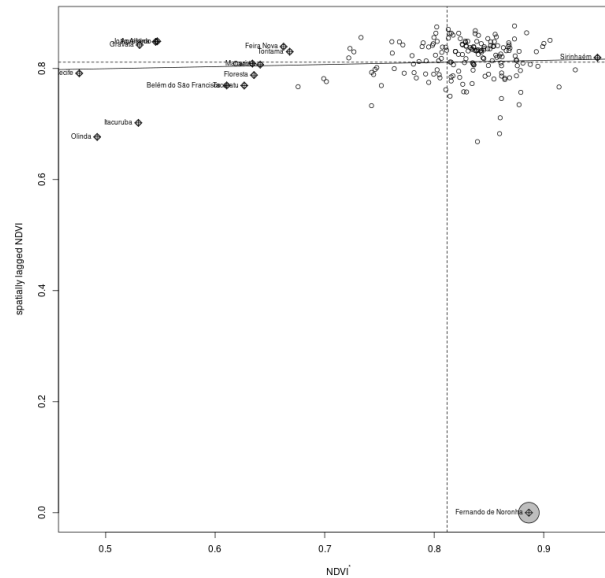
ao considerar nível de 1% correlação espacial nos dados. Uma visualização gráfica pode ser observada na Figura 6a.

A Figura 6b apresenta o LISA (Local Indicators of Spatial Association) e a autocorrelação local dos índices de vegetação NDVI* e EVI nos municípios analisados.

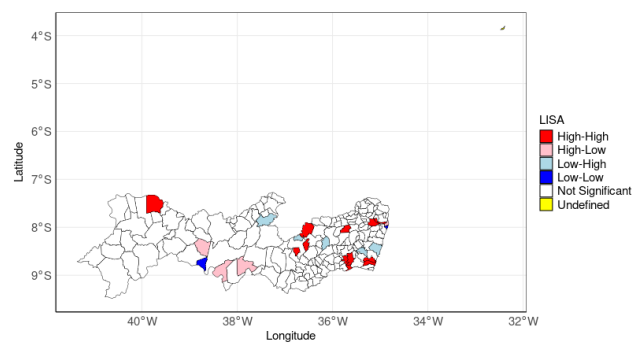
A análise espacial por meio do índice LISA Figura 6b revela agrupamentos significativos dos índices de vegetação NDVI* e EVI nos municípios analisados. Observam-se clusters High-High (vermelho), indicando áreas com alta vegetação cercadas por municípios igualmente vegetados, e Low-Low (azul escuro), associados a regiões com possível degradação ambiental. Outliers espaciais do tipo High-Low (rosa) e Low-High (azul claro) sugerem transições locais na cobertura vegetal, possivelmente resultantes de pressões antrópicas ou alterações no uso do solo. Municípios classificados como Not Significant (branco) não apresentaram autocorrelação estatisticamente relevante, enquanto os em Undefined (amarelo) não tiveram padrão claro.

Dados com autocorrelação espacial exigem métodos específicos para lidar com essa dependência. Nossa proposta, a regressão GOLL- $\mathcal{T}$, necessita que os dados sejam independentes; assim, foi analisado o índice de Moran Global e o LISA (Local Indicators of Spatial Association) para selecionar valores de NDVI* e EVI que não possuíam dependência espacial. O NDVI selecionado foram os valores acima do quantil 0.99, ou seja, com foco nas regiões de vegetação

densa, como florestas tropicais e matas nativas, onde a cobertura do solo é quase total. Para o EVI, foi adotada a mediana.



(a) Análise da autocorrelação espacial com Índice de Moran Global



(b) Índice de Moran Local (LISA) para o estado de Pernambuco

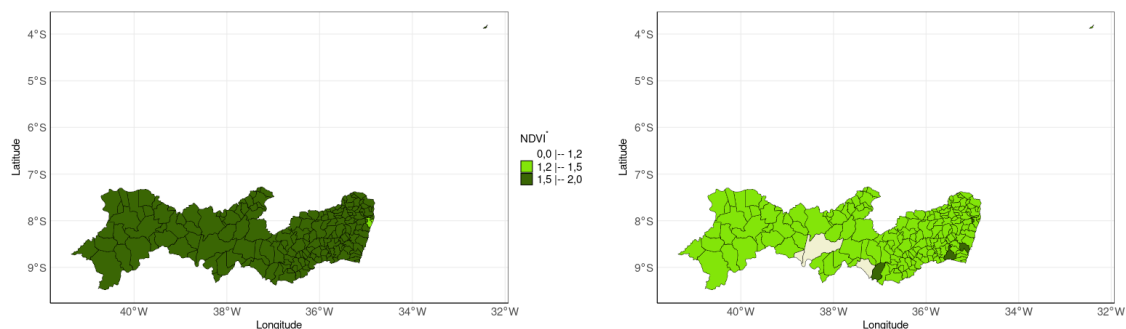
Figura 6 – Autocorrelação espacial com Índice de Moran Global e mapa de Pernambuco com informações do índice de Moran Local (LISA).

4.2.4 Mapas de Pernambuco - NDVI e EVI

A Figura 7a ilustra a distribuição espacial do NDVI nos municípios de Pernambuco, destacando as variações na cobertura vegetal ao longo do estado. Observa-se que a maior

parte do território estadual apresenta valores de NDVI moderados a altos, indicando áreas com vegetação densa ou bem preservada. No entanto, algumas regiões isoladas ao leste do estado exibem valores mais baixos, o que pode ser atribuído a fatores como diferenças climáticas locais, mudanças no uso da terra, como a urbanização ou a agricultura intensiva, ou ainda a condições específicas de solo que limitam o crescimento vegetal.

A Figura 7 ilustra a distribuição espacial do EVI nos municípios de Pernambuco, proporcionando uma avaliação refinada da biomassa vegetal ao reduzir os efeitos de interferências atmosféricas e variabilidades do solo. As áreas representadas por tonalidades mais escuras correspondem a regiões com maior densidade de vegetação, com valores de EVI variando entre 0,5 e 1,0, enquanto as tonalidades mais claras indicam regiões com cobertura vegetal mais escassa, com EVI variando entre 0,0 e 0,2. É possível observar que grande parte do estado apresenta valores moderados a baixos de EVI, sugerindo uma predominância de vegetação menos densa ou impactada. No entanto, algumas áreas pontuais, com valores mais altos de EVI, destacam-se como regiões de vegetação mais densa, possivelmente associadas a remanescentes florestais ou áreas menos afetadas pela conversão para usos agrícolas ou urbanos, refletindo a presença de ecossistemas naturais ou áreas de preservação ambiental.



(a) Mapa de Pernambuco - NDVI*

(b) Mapa Pernambuco - EVI

Figura 7 – Mapas de Pernambuco com informações de NDVI* e EVI

4.2.5 Análise de regressão

Na Tabela 5 são apresentadas estimativas dos parâmetros, erro padrão correspondente, estatística de teste t e p-valor. Note que, o intercepto e a covariável explicativa são estatisticamente

significativas ao nível nominal de pelo menos 1%.

Tabela 5 – Estimação pontual, erro padrão, estatística de teste t e P-valor

Parâmetros	Estimativa	erro padrão	estatística t	p-valor
β_0	1.3843	0.2936	4.7156	4.7641×10^{-6}
β_1	2.8024	0.8479	3.3445	1.0001×10^{-3}

O EVI e o NDVI* foram modelados por meio da regressão, tendo o EVI como variável explicativa. A variação do NDVI* em resposta a mudanças no EVI pode ser descrita pela fórmula:

$$\frac{\mu_i^*}{\mu_i} = e^{\beta_1} \cdot \frac{1 + \eta_i}{1 + e^{\beta_1} \cdot \eta_i},$$

em que, $\eta_i = e^{\beta_0} e^{\beta_1 x_i}$, $\mu_i = E\{Y_i|X_i = x_i\} = \frac{2e^{\beta_0} e^{\beta_1 x_i}}{1 + e^{\beta_0} e^{\beta_1 x_i}}$, e $\mu_i^* = E\{Y_i|X_i = x_i + 1\} = \frac{2e^{\beta_0} e^{\beta_1 x_i}}{1 + e^{\beta_0} e^{\beta_1 x_i}}$. Essa relação permite avaliar o efeito proporcional da variação no EVI sobre o NDVI*, considerando o estado atual da vegetação, representado por η_i . Por exemplo, adotando-se um valor hipotético $\eta_i = 3$, observa-se que um incremento unitário em EVI resulta em um aumento de aproximadamente 30% no NDVI*.

A Figura 8 apresenta o ajuste dos modelos de regressão GOLL- $\mathcal{T}$ e modelo de regressão linear aos dados observados, permitindo uma visualização comparativa das abordagens. Os pontos sob a variável resposta NDVI* se concentram em uma faixa estreita de valores, com a maioria das observações situadas entre 1,5 e 2,0. Além disso, a variável preditora EVI está distribuída em um intervalo restrito, concentrando-se em valores menores que 0,5. Esse padrão sugere que a relação entre as variáveis pode não ser linear, justificando a aplicação de modelos mais flexíveis.

O modelo GOLL- $\mathcal{T}$ apresenta maior capacidade de adaptação às variações da variável resposta, capturando melhor a tendência dos dados, especialmente nas regiões onde a quantidade de observações é mais alta. Já o modelo linear impõe uma relação fixa entre as variáveis, resultando em um ajuste menos preciso em áreas onde a resposta apresenta variações não lineares.

Através da Figura 8, é possível perceber que os dados não seguem um padrão linear. Esse padrão reforça a hipótese de que a relação entre as variáveis não é bem representada por uma regressão linear simples, sendo necessário um modelo mais flexível para capturar adequadamente a estrutura dos dados.

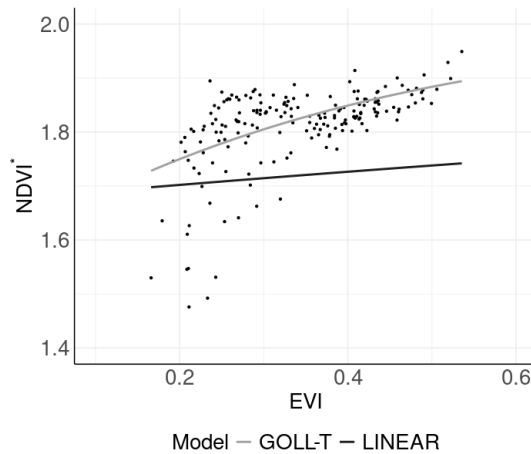
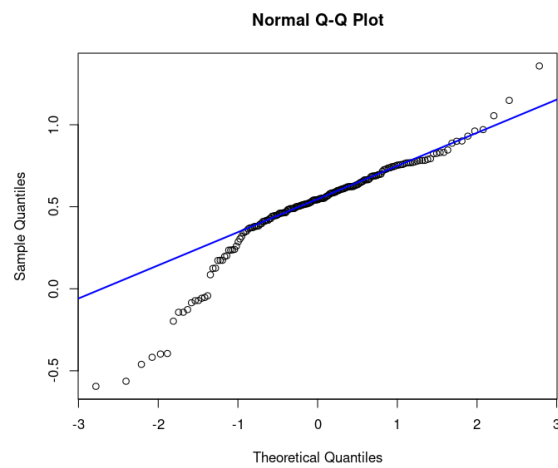


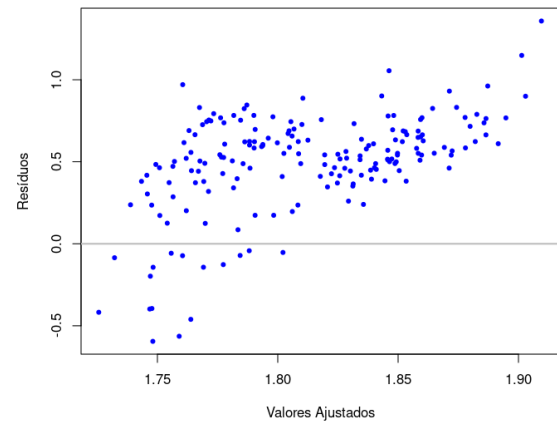
Figura 8 – Ajuste do modelos de regressão GOLL- $\mathcal{T}$ e modelo de regressão linear.

A Figura 9a avalia a normalidade dos resíduos do modelo ajustado. Observe-se que, na maior parte da distribuição, especialmente para os quantis intermediários, os pontos seguem aproximadamente a linha teórica, indicando um bom ajuste à normalidade. No entanto, há desvios nas extremidades, especialmente nas caudas inferior e superior, indicando a presença de resíduos com distribuição mais pesada (caudas longas) do que a normal. Isso pode indicar a necessidade de um modelo mais flexível para capturar essas discrepâncias. Já a Figura 9b mostra o gráfico dos resíduos em relação aos valores ajustados. Observa-se um padrão de curvatura crescente, indicando uma possível violação da suposição de linearidade. Além disso, há indícios de heterocedasticidade, uma vez que a variância dos resíduos aumenta à medida que os valores ajustados crescem. Esses comportamentos sugerem que o modelo linear pode não ser o mais adequado para representar a relação entre as variáveis, sendo recomendada a consideração de transformações ou modelos mais flexíveis.

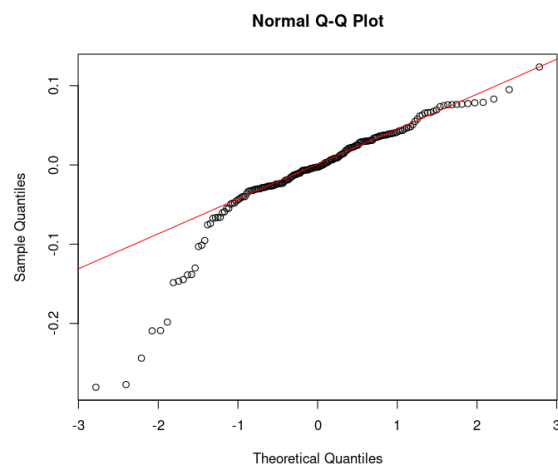
A Figura 9c apresenta o gráfico Q-Q dos resíduos do modelo GOLL- $\mathcal{T}$. Observa-se que, para uma ampla faixa dos quantis centrais, os pontos se alinham bem à linha teórica da normal, indicando que os resíduos seguem um padrão aproximadamente normal nessa região. No entanto, são percebidos desvios nas extremidades, com maior destaque para a cauda inferior, sugerindo assimetria negativa leve ou presença de valores extremos. Ainda assim, o comportamento global dos resíduos aponta para uma aproximação razoável da normalidade, especialmente quando comparado ao modelo linear. A Figura 9d exibe o gráfico de resíduos versus valores ajustados para o modelo GOLL- $\mathcal{T}$. Os resíduos apresentam distribuição aproximadamente simétrica em torno de zero, sem evidência de padrões sistemáticos ou tendência clara, o que sugere um bom ajuste do modelo. A variabilidade dos resíduos é relativamente constante ao longo da faixa de valores ajustados, indicando atendimento da suposição de homocedasticidade. Quando



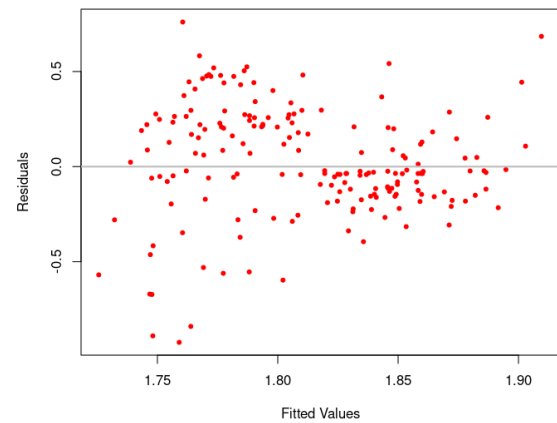
(a) QQ-plot Regressão Linear Múltipla



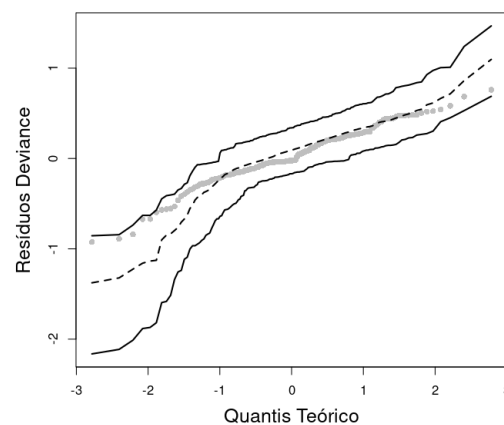
(b) Resíduos vs Valores Ajustados - Linear Múltiplo



(c) QQ-plot GOLL-T



(d) Resíduos vs Valores Ajustados - GOLL-T



(e) Resíduo Deviance

Figura 9 – Análise dos Resíduos

comparado ao modelo linear, observa-se uma dispersão mais homogênea dos resíduos, o que reforça a adequação superior do modelo GOLL- $\mathcal{T}$ na captura das características dos dados.

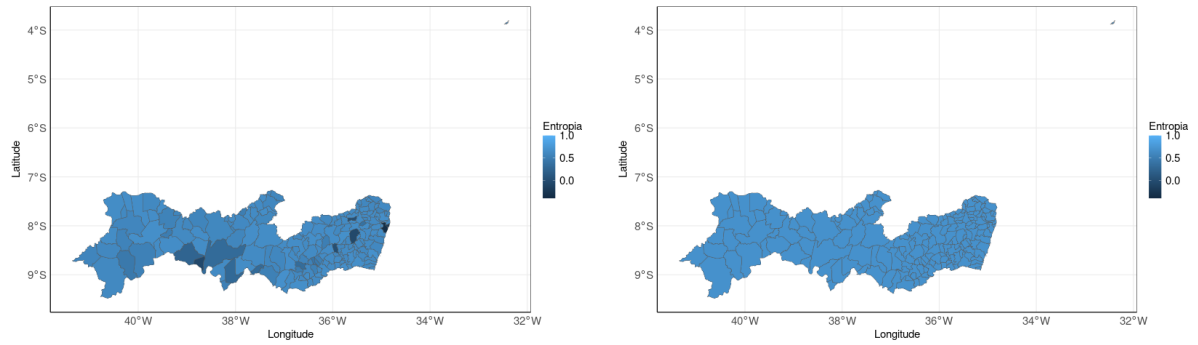
A Figura 9e apresenta o gráfico quantil-quantil (QQ-plot) dos resíduos deviance, permitindo avaliar a adequação do modelo ajustado à distribuição dos dados. Os resíduos deviance seguem aproximadamente a linha dos quantis teóricos normais, indicando que o modelo ajustado captura bem a estrutura geral dos dados. Pequenos desvios são esperados, especialmente em amostras finitas, e não comprometem a validade do modelo. Observa-se que, em algumas regiões das caudas, há um leve afastamento dos resíduos em relação à linha de referência. Contudo, o modelo mantém uma boa aderência na região central, que concentra a maioria das observações. A maioria dos resíduos permanece dentro das bandas de confiança representadas pelas linhas sólida e pontilhada, reforçando a adequação do modelo. Pequenos desvios dentro dessas bandas não indicam necessariamente um problema estrutural, mas sim a natural variabilidade dos dados. A proximidade dos resíduos à linha de referência sugere que o modelo GOLL- $\mathcal{T}$ fornece uma representação adequada aos dados.

A Figura 10a A entropia observada foi utilizada como uma medida de desordem ecológica associada à diversidade de vegetação no estado de Pernambuco. De modo geral, o estado apresenta índices elevados de vegetação, e essa variação está diretamente relacionada ao valor da entropia: áreas com maior diversidade de tipos vegetais tendem a apresentar valores mais altos de entropia. A interpretação gráfica da entropia permite identificar que, em regiões onde há uma mistura mais heterogênea de vegetação ou seja, diferentes tipos vegetais distribuídos no espaço, a entropia tende a ser mais elevada. Por outro lado, áreas com tons mais escuros no mapa indicam valores mais baixos de entropia, sugerindo menor diversidade de vegetação. Nessas regiões, a cobertura vegetal é ausente ou composta predominantemente por um único tipo de vegetação. A região do Recife, por exemplo, apresentou um valor reduzido de entropia, o que pode ser explicado pela menor variabilidade na cobertura vegetal, característica comum em áreas urbanizadas ou densamente ocupadas. Assim, a entropia observada fornece uma representação espacial da heterogeneidade vegetal, sendo útil como ferramenta interpretativa da dinâmica ecológica em diferentes localidades do estado.

A Figura 10b no que se refere à entropia predita, o modelo estimado considerou a influência do EVI, estabelecendo um intervalo fixo entre 0 e 1. Os resultados demonstraram que a entropia estimada apresenta padrão semelhante ao da entropia observada, o que sugere uma boa capacidade preditiva do modelo. Isso reforça que o EVI, ao ser utilizado como covariável, consegue explicar de forma satisfatória a distribuição espacial da entropia vegetal em Pernambuco.

As Figuras 10a e 10b são mapas de Pernambuco utilizando entropia observada e entropia

predita. Para criação desse mapa foi utilizado a equação 2.12, em que a entropia observada foi utilizado o verdadeiro valor do NDVI na equação, e a entropia predita foi utilizado os valores do NDVI preditos pelo modelo GOLL- $\mathfrak{T}$.



(a) Mapa de Pernambuco - Entropia observada

(b) Mapa Pernambuco - Entropia predita

Figura 10 – Mapas de Pernambuco com valores da entropia observada e predita para cada município

5 CONCLUSÃO

Esta dissertação objetivou a modelagem de fenômenos complexos, mostrando que distribuições mais flexíveis, como a GOLL- $\mathcal{T}$, oferecem bons ajustes e de fácil interpretação, especialmente em cenários com relações não lineares.

Especificamente, desenvolveu-se uma nova distribuição reparametrizada em torno da média com base na família GOLL-G. Foram apresentadas algumas propriedades dessa distribuição, tais como variância, curtose, assimetria e entropia de Shannon. Além disso, foi proposto o modelo de regressão GOLL- $\mathcal{T}$ e, para tal, derivou-se algumas propriedades analíticas associadas, como as versões matriciais da função *score* e da matriz de informação de Fisher. Também o estimador *scoring*-Fisher para os parâmetros da regressão proposta foi desenvolvido. Sob uma perspectiva numérica, realizou-se um estudo de simulação Monte Carlo com o objetivo de quantificar o desempenho dos estimadores de máxima verossimilhança em diferentes cenários. Os resultados mostraram que, à medida que o tamanho da amostra n aumenta, o viés e o erro quadrático médio (EQM) dos estimadores diminuem, confirmando o comportamento assintótico esperado. Por fim, aplicou-se o modelo de regressão GOLL- $\mathcal{T}$ para avaliar o impacto da variável EVI sobre a média do NDVI. Os resultados evidenciam que modelos mais flexíveis, como o GOLL- $\mathcal{T}$, capturam melhor a complexidade da relação entre as variáveis, superando o modelo linear em termos de ajuste e capacidade preditiva.

Como perspectiva futura, seria interessante explorar a reformulação do modelo proposto para dados espaço temporais, permitindo analisar como o índice de vegetação e outras variáveis ambientais se relacionam e evoluem ao longo do tempo.

REFERÊNCIAS

- ACCIOLY, L. J. et al. Relações empíricas entre a estrutura da vegetação e dados do sensor tm/landsat. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 6, n. 3, p. 492–498, 2002.
- ALZAATREH, A.; LEE, C.; FAMOYE, F. A new method for generating families of continuous distributions. *Metron*, Springer, v. 71, n. 1, p. 63–79, 2013.
- BALAGHI, R.; TYCHON, B.; EERENS, H.; JLIBENE, M. Empirical regression models using ndvi, rainfall and temperature data for the early prediction of wheat grain yields in morocco. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 10, n. 4, p. 438–452, 2008.
- CABRERA-BOSQUET, L.; MOLERO, G.; STELLACCI, A.; BORT, J.; NOGUÉS, S.; ARAUS, J. Ndzi as a potential tool for predicting biomass, plant nitrogen content and growth in wheat genotypes subjected to different water and nitrogen conditions. *Cereal Research Communications*, Akadémiai Kiadó, v. 39, n. 1, p. 147–159, 2011.
- CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2. ed. Pacific Grove, CA: Duxbury, 2002.
- CHAMBERS, J. M. *Software for data analysis: programming with R*. [S.l.]: Springer, 2008. v. 2.
- CORDEIRO, A. P. A.; BERLATO, M. A.; FONTANA, D. C.; MELO, R. W. d.; SHIMABUKURO, Y. E.; FIOR, C. S. Regiões homogêneas de vegetação utilizando a variabilidade do ndvi. *Ciência Florestal*, SciELO Brasil, v. 27, p. 883–896, 2017.
- CORDEIRO, G. M.; ALIZADEH, M.; OZEL, G.; HOSSEINI, B.; ORTEGA, E. M. M.; ALTUN, E. The generalized odd log-logistic family of distributions: properties, regression models and applications. *Journal of statistical computation and simulation*, Taylor & Francis, v. 87, n. 5, p. 908–932, 2017.
- DAMIAN, J. M.; PIAS, O. H. d. C.; CHERUBIN, M. R.; FONSECA, A. Z. d.; FORNARI, E. Z.; SANTI, A. L. Applying the ndvi from satellite images in delimiting management zones for annual crops. *Scientia Agrícola*, SciELO Brasil, v. 77, n. 1, p. e20180055, 2019.
- FITZGERALD, D. Analysis of extreme rainfall using the log logistic distribution. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 19, p. 249–257, 2005.
- FOODY, G. Geographical weighting as a further refinement to regression modelling: An example focused on the ndvi–rainfall relationship. *Remote sensing of Environment*, Elsevier, v. 88, n. 3, p. 283–293, 2003.
- FUNG, T.; SIU, W. Environmental quality and its changes, an analysis using ndvi. *International Journal of Remote Sensing*, Taylor & Francis, v. 21, n. 5, p. 1011–1024, 2000.
- GLEATON, J.; LYNCH, J. Properties of generalized log-logistic families of lifetime distributions. *Journal of Probability and Statistical Science*, v. 4, n. 1, p. 51–64, 2006.
- GLEATON, J. U.; LYNCH, J. D. Extended generalized log-logistic families of lifetime distributions with an application. *J. Probab. Stat. Sci*, v. 8, n. 1, p. 1–17, 2010.

- GLENN, E. P.; NEALE, C. M.; HUNSAKER, D. J.; NAGLER, P. L. Vegetation index-based crop coefficients to estimate evapotranspiration by remote sensing in agricultural and natural ecosystems. *Hydrological Processes*, Wiley Online Library, v. 25, n. 26, p. 4050–4062, 2011.
- GUPTA, R. C.; GUPTA, R. D. Proportional reversed hazard rate model and its applications. *Journal of statistical planning and inference*, Elsevier, v. 137, n. 11, p. 3525–3536, 2007.
- HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote sensing of environment*, Elsevier, v. 83, n. 1-2, p. 195–213, 2002.
- HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote Sensing of Environment*, v. 83, n. 1-2, p. 195–213, 2002.
- HUETE, A.; JUSTICE, C.; LIU, H. Development of vegetation and soil indices for modis-eos. *Remote Sensing of environment*, Elsevier, v. 49, n. 3, p. 224–234, 1994.
- HUETE, A.; LIU, H.; BATCHILY, K.; LEEUWEN, W. V. A comparison of vegetation indices over a global set of tm images for eos-modis. *Remote sensing of environment*, Elsevier, v. 59, n. 3, p. 440–451, 1997.
- KARIYEVA, J.; LEEUWEN, W. J. D. V. Environmental drivers of ndvi-based vegetation phenology in central asia. *Remote Sensing*, v. 3, n. 2, p. 203–246, 2011.
- KURBANOV, R. K.; ZAKHAROVA, N. I. Application of vegetation indexes to assess the condition of crops. *Agricultural Machinery and Technologies*, v. 14, n. 4, p. 4–11, 2020.
- LEE, S. Application of logistic regression model and its validation for landslide susceptibility mapping using gis and remote sensing data. *International Journal of remote sensing*, Taylor & Francis, v. 26, n. 7, p. 1477–1491, 2005.
- LI, Z.; LI, X.; WEI, D.; XU, X.; WANG, H. An assessment of correlation on modis-ndvi and evi with natural vegetation coverage in northern hebei province, china. *Procedia Environmental Sciences*, Elsevier, v. 2, p. 964–969, 2010.
- LIMA, M. d. C. S.; PRATAVIERA, F.; ORTEGA, E. M.; CORDEIRO, G. M. The odd log-logistic geometric family with applications in regression models with varying dispersion. *Journal of Statistical Theory and Applications*, Springer, v. 18, n. 3, p. 278–294, 2019.
- MAGALHÃES, M. N.; LIMA, A. C. P. d. *Noções de Probabilidade e Estatística*. 7. ed.. ed. São Paulo: Editora da Universidade de São Paulo, 2011. ISBN 978-85-314-0677-5.
- PAULA, G. A. *Modelos de Regressão com Apoio Computacional*. 2004. Instituto de Matemática e Estatística, Universidade de São Paulo. E-mail: giapaula@ime.usp.br.
- PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M. *Sensoriamento remoto da vegetação*. [S.l.]: Oficina de textos, 2015.
- PROPASTIN, P. A.; KAPPAS, M. Reducing uncertainty in modeling the ndvi-precipitation relationship: a comparative study using global and local regression techniques. *GIScience & Remote Sensing*, Taylor & Francis, v. 45, n. 1, p. 47–67, 2008.

- RAHMAN, M. R.; ISLAM, A.; RAHMAN, M. A. Ndzi derived sugarcane area identification and crop condition assessment. *Plan Plus*, v. 1, n. 2, p. 1–12, 2004.
- RISTIČ, M. M.; BALAKRISHNAN, N. A distribuição exponencial gama-exponencial. *Jornal de computação e simulação estatística*, Taylor & Francis, v. 82, n. 8, p. 1191–1206, 2012.
- ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the great plains with erts. In: *NASA Goddard Space Flight Center 3D ERTS-1Symp.* [S.l.: s.n.], 1974. v. 1, p. 309–317.
- SHANNON, C. E. A mathematical theory of communication. *The Bell system technical journal*, Nokia Bell Labs, v. 27, n. 3, p. 379–423, 1948.
- SILVA, V. S. da; SALAMI, G.; SILVA, M. I. O. da; SILVA, E. A.; JUNIOR, J. J. M.; ALBA, E. Methodological evaluation of vegetation indexes in land use and land cover (lulc) classification. *Geology, Ecology, and Landscapes*, Taylor & Francis, v. 4, n. 2, p. 159–169, 2020.
- TAHIR, M.; CORDEIRO, G. M.; ALZAATREH, A.; MANSOOR, M.; ZUBAIR, M. The logistic-x family of distributions and its applications. *Communications in statistics-Theory and methods*, Taylor & Francis, v. 45, n. 24, p. 7326–7349, 2016.
- TENREIRO, T. R.; GARCÍA-VILA, M.; GÓMEZ, J. A.; JIMÉNEZ-BERNI, J. A.; FERERES, E. Using ndvi for the assessment of canopy cover in agricultural crops within modelling research. *Computers and Electronics in Agriculture*, Elsevier, v. 182, p. 106038, 2021.
- VIGAS, V. P.; ORTEGA, E. M.; SUZUKI, A. K.; CORDEIRO, G. M.; JUNIOR, P. C. dos S. The generalized odd log-logistic-g regression with interval-censored survival data. *Journal of Applied Statistics*, Taylor & Francis, v. 51, n. 9, p. 1642–1663, 2024.
- WELCH, R. M. et al. Impact of deforestation on the proposed mesoamerican biological corridor in central america. In: *CONF ON HYDROLOGY, 19., AMS ANNUAL MEETING, 85.* [S.l.: s.n.], 2005. Anais...
- ZANZARINI, F. V.; PISSARRA, T. C.; BRANDÃO, F. J.; TEIXEIRA, D. D. Correlação espacial do índice de vegetação (ndvi) de imagem landsat/etm+ com atributos do solo. *Revista Brasileira de Engenharia Agrícola e Ambiental*, SciELO Brasil, v. 17, p. 608–614, 2013.
- ZHU, X.; LIU, D. Improving forest aboveground biomass estimation using seasonal landsat ndvi time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 102, p. 222–231, 2015.
- ZOGRAFOS, K.; BALAKRISHNAN, N. Sobre famílias de distribuições beta e gama generalizadas geradas e inferência associada. *Metodologia estatística*, Elsevier, v. 6, n. 4, p. 344–362, 2009.

APÊNDICE A – DEDUÇÃO MATEMÁTICA DA DISTRIBUIÇÃO E PROPRIEDADES ASSOCIADAS

A.1 DISTRIBUIÇÃO E FUNÇÃO DENSIDADE DE PROBABILIDADE DE GOLL- $\mathfrak{T}$

Seja Y uma variável aleatória com função densidade de probabilidade $f(y)$ dada por:

$$f(y) = \frac{\mu}{2 - \mu} 2^{-\frac{\mu}{2-\mu}} y^{\frac{\mu}{2-\mu}-1} \mathbb{I}_{(0,2)}.$$

A função de distribuição acumulada (fda) pode ser obtida pela seguinte dedução:

$$\begin{aligned} F(y) &= \int_0^y \frac{\mu}{2 - \mu} 2^{-\frac{\mu}{2-\mu}} z^{\frac{\mu}{2-\mu}-1} dz \\ &= \frac{\mu}{2 - \mu} 2^{-\frac{\mu}{2-\mu}} \int_0^y z^{\frac{\mu}{2-\mu}-1} dz \\ &= \frac{\mu}{2 - \mu} 2^{-\frac{\mu}{2-\mu}} \frac{z^{\frac{\mu}{2-\mu}-1+1}}{\frac{\mu}{2-\mu}-1+1} \Big|_0^y \\ &= \frac{\mu}{2 - \mu} 2^{-\frac{\mu}{2-\mu}} \frac{y^{\frac{\mu}{2-\mu}}}{\frac{\mu}{2-\mu}} \\ &= 2^{-\frac{\mu}{2-\mu}} y^{\frac{\mu}{2-\mu}}. \end{aligned}$$

A.2 PROPRIEDADES DA DISTRIBUIÇÃO GOLL- $\mathfrak{T}$

Algumas propriedades da distribuição como variância, assimetria, curtose foram calculadas. A seguir serão descritos os quatros primeiros momentos da distribuição GOLL- $\mathfrak{T}$ e as propriedades mencionadas anteriormente.

Primeiro momento é dado por: $E(Y) = \mu$.

Segundo momento é dado por $E(Y^2)$:

$$\begin{aligned}
E(Y^2) &= \int_0^2 y^2 f(y) dy \\
&= \int_0^2 y^2 \frac{\mu}{2-\mu} \frac{1}{2^{\frac{\mu}{2-\mu}}} y^{\frac{\mu}{2-\mu}-1} dy \\
&= \frac{\mu}{2-\mu} \frac{1}{2^{\frac{\mu}{2-\mu}}} \int_0^2 y^2 y^{\frac{\mu}{2-\mu}-1} dy \\
&= \frac{\mu}{2-\mu} \frac{1}{2^{\frac{\mu}{2-\mu}}} \int_0^2 y^{\frac{\mu}{2-\mu}-1+2} dy \\
&= \frac{\mu}{2-\mu} \frac{1}{2^{\frac{\mu}{2-\mu}}} \frac{y^{\frac{\mu}{2-\mu}+2}}{\frac{\mu}{2-\mu}+2} \Big|_0^2 \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \frac{2^{\frac{\mu}{2-\mu}+2} (2-\mu)}{\mu+4-2\mu} \\
&= \frac{\mu}{2-\mu} \frac{2^{-\frac{\mu}{2-\mu}+\frac{\mu}{2-\mu}+2} (2-\mu)}{\mu+4-2\mu} \\
&= \frac{\mu}{2-\mu} \frac{4(2-\mu)}{4-\mu} \\
&= \frac{4\mu}{4-\mu}.
\end{aligned}$$

Terceiro momento é dado por $E(Y^3)$:

$$\begin{aligned}
E(Y^3) &= \int_0^2 y^3 f(x) dy \\
&= \int_0^2 y^3 \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} y^{\frac{\mu}{2-\mu}-1} dy \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \int_0^2 y^3 y^{\frac{\mu}{2-\mu}-1} dy \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \int_0^2 y^{\frac{\mu}{2-\mu}-1+3} dy \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \frac{y^{\frac{\mu}{2-\mu}+3}}{\frac{\mu}{2-\mu}+3} \Big|_0^2 \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \frac{2^{\frac{\mu}{2-\mu}+3}}{\frac{\mu+3(2-\mu)}{2-\mu}} \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \frac{2^{\frac{\mu}{2-\mu}+3} (2-\mu)}{\mu+6-3\mu} \\
&= \frac{\mu}{2-\mu} \frac{2^{-\frac{\mu}{2-\mu}} 2^{\frac{\mu}{2-\mu}+3} (2-\mu)}{6-2\mu} \\
&= \frac{8\mu}{6-2\mu} \\
&= \frac{4\mu}{3-\mu}.
\end{aligned}$$

Quarto momento é dado por $E(Y^4)$:

$$\begin{aligned}
E(Y^4) &= \int_0^2 y^4 f(y) dy \\
&= \int_0^2 y^4 \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} y^{\frac{\mu}{2-\mu}-1} dy \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \int_0^2 y^4 y^{\frac{\mu}{2-\mu}-1} dy \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \int_0^2 x^{\frac{\mu}{2-\mu}-1+4} dy \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \left. \frac{y^{\frac{\mu}{2-\mu}+4}}{\frac{\mu}{2-\mu}+4} \right|_0^2 \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \frac{2^{\frac{\mu}{2-\mu}+4}}{\frac{\mu+4(2-\mu)}{2-\mu}} \\
&= \frac{\mu}{2-\mu} 2^{-\frac{\mu}{2-\mu}} \frac{2^{\frac{\mu}{2-\mu}+4} (2-\mu)}{\mu+8-4\mu} \\
&= \frac{\mu}{2-\mu} \frac{2^{-\frac{\mu}{2-\mu}} 2^{\frac{\mu}{2-\mu}+4} (2-\mu)}{\mu+8-4\mu} \\
&= \frac{16\mu}{8-3\mu}.
\end{aligned}$$

O valor esperado do $\log(Y)$ é dado por, para $c = \frac{\mu}{2-\mu}$

$$\begin{aligned}
E[\log(Y)] &= \int_0^2 \log(y) \frac{c}{2^c} y^{c-1} dy = 1. \\
&= \int_0^2 \log(y) y^{c-1} dy = \frac{2^c}{c} \\
&= \int_0^2 (\log y) y^{c-1} dy = \frac{\partial}{\partial c} \left[\frac{2^c}{c} \right] \\
&= \int_0^2 \log(y) \frac{c}{2^c} y^{c-1} dy = \frac{c}{2^c} \frac{\partial}{\partial c} \left[\frac{2^c}{c} \right] \\
&= \frac{c}{2^c} \left[\frac{c(\log 2) 2^c - 2^c}{c^2} \right] \\
&= \frac{2^c c}{2^c c} [c(\log 2) - 1] \\
&= \frac{(\log 2)c - 1}{c} \\
&= \log(2) - \frac{1}{c} \\
&= -\frac{1}{\left(\frac{\mu}{2-\mu}\right)} + \log(2) \\
&= -\frac{(2-\mu)}{\mu} + \log(2).
\end{aligned}$$

A.2.1 Variância

A variância é dada por:

$$\begin{aligned}
 Var(Y) &= E(Y^2) - [E(Y)]^2 \\
 &= \frac{4\mu}{4-\mu} - \mu^2 \\
 &= \frac{4\mu - \mu^2(4-\mu)}{4-\mu} \\
 &= \frac{4\mu - 4\mu^2 + \mu^3}{4-\mu} \\
 &= \frac{\mu(4 - 4\mu + \mu^2)}{4-\mu} \\
 &= \frac{\mu(2-\mu)^2}{4-\mu}.
 \end{aligned}$$

A.2.2 Assimetria

Seja Y uma variável aleatória com valor esperado μ e o desvio padrão σ . O coeficiente de assimetria de Y , denotado por γ_3 , indica o grau de assimetria da sua distribuição de probabilidade e é definido por

$$\gamma_3 = \frac{E(Y^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}, \quad (A.1)$$

em que supomos a existência do terceiro momento de Y (MAGALHÃES; LIMA, 2011).

Para encontrar a assimetria, precisamos do valor de σ^3 :

$$\sigma^3 = \left(\sqrt{\frac{\mu(2-\mu)^2}{4-\mu}} \right)^3 = \left(\frac{\mu(2-\mu)^2}{4-\mu} \right)^{\frac{3}{2}} = \frac{\mu^{\frac{3}{2}}(2-\mu)^3}{(4-\mu)^{\frac{3}{2}}}.$$

Utilizando (A.1), temos que a assimetria é

$$\begin{aligned}
\gamma_3 &= \frac{E(Y^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3} \\
&= \frac{\frac{4\mu}{3-\mu} - 3\mu \left(\frac{\mu(2-\mu)^2}{(4-\mu)} \right) - \mu^3}{\frac{\mu^{\frac{3}{2}}(2-\mu)^3}{(4-\mu)^{\frac{3}{2}}}} \\
&= \frac{\frac{4\mu}{(3-\mu)} - \frac{3\mu^2(4-4\mu+\mu^2)}{4-\mu} - \mu^3}{\frac{\mu^{\frac{3}{2}}(2-\mu)^3}{(4-\mu)^{\frac{3}{2}}}} \\
&= \frac{\frac{4\mu}{3-\mu} - \frac{12\mu^2-12\mu^3+3\mu^4}{4-\mu} - \mu^3}{\frac{\mu^{\frac{3}{2}}(2-\mu)^3}{(4-\mu)^{\frac{3}{2}}}}.
\end{aligned}$$

A.2.3 Curtose

O coeficiente de curtose da variável aleatória Y , denotado por γ_4 , é definido por

$$\gamma_4 = \frac{E[(X - \mu)^4]}{\sigma^4},$$

em que supomos a existência do quarto momento de Y (MAGALHÃES; LIMA, 2011). Para nosso caso,

$$\sigma^4 = (\sigma^2)^2 = \left(\frac{\mu(2-\mu)^2}{(4-\mu)} \right)^2 = \frac{\mu^2(2-\mu)^4}{(4-\mu)^2}. \quad (\text{A.2})$$

Agora, calculando $E[(Y - \mu)^4]$, que representa o quarto momento central da variável aleatória Y ,

$$\begin{aligned}
E(Y - \mu)^4 &= EY^4 - 4EY^3\mu + 6EY^2\mu^2 - 4EY\mu^3 + \mu^4 \\
&= EY^4 - 4EY^3\mu + 6\mu^2(\mu^2 + \sigma^2) - 4\mu^4 + \mu^4 \\
&= EY^4 - 4EY^3\mu + 6\mu^4 + 6\mu^2\sigma^2 - 4\mu^4 + \mu^4 \\
&= EY^4 - 4EY^3\mu + 3\mu^4 + 6\mu^2\sigma^2.
\end{aligned} \quad (\text{A.3})$$

Assim, é possível obter a curtose como:

$$\begin{aligned}
\gamma_4 &= \frac{E[(Y - \mu)^4]}{\sigma^4} \\
&= \frac{\frac{16\mu}{8-3\mu} - 4\mu\left(\frac{4\mu}{3-\mu}\right) + 3\mu^4 + 6\mu^2\left[\frac{\mu(2-\mu)^2}{4-\mu}\right]}{\frac{\mu^2(2-\mu)^4}{(4-\mu)^2}} \\
&= \frac{\frac{16\mu}{8-3\mu} - \frac{16\mu^2}{3-\mu} + 3\mu^4 + \frac{6\mu^3(2-\mu)^2}{4-\mu}}{\frac{\mu^2(2-\mu)^4}{(4-\mu)^2}}
\end{aligned}$$

A.2.4 Entropia de Shannon

A seguir pode ser encontrado a entropia de Shannon para a função densidade GOLL- $\mathfrak{T}$. A entropia de Shannon é obtida como $E[-\log f(Y)]$:

$$\begin{aligned}
E[-\log f(y)] &= \int_0^2 -\log f(y) f(y) dy \\
&= \int_0^2 - \left[\log \frac{\mu}{2-\mu} - \frac{\mu}{2-\mu} \log(2) + \left(\frac{\mu}{2-\mu} - 1 \right) \log(y) \right] f(y) dy \\
&= \int_0^2 - \left[\log \frac{\mu}{2-\mu} f(y) - \frac{\mu}{2-\mu} \log(2) f(y) \right. \\
&\quad \left. + \left(\frac{\mu}{2-\mu} - 1 \right) \log(y) f(y) \right] dy \\
&= - \left\{ \int_0^2 \log \frac{\mu}{2-\mu} f(y) dy - \int_0^2 \frac{\mu}{2-\mu} \log(2) f(y) dy \right. \\
&\quad \left. + \int_0^2 \left(\frac{\mu}{2-\mu} - 1 \right) \log(y) f(y) dy \int_0^2 \log(y) f(y) dy \right\} \\
&= - \left\{ \log \frac{\mu}{2-\mu} \int_0^2 f(y) dy - \frac{\mu}{2-\mu} \log(2) \int_0^2 f(y) dy \right. \\
&\quad \left. + \left(\frac{\mu}{2-\mu} - 1 \right) \int_0^2 \log(y) f(y) dy \right\} \tag{A.4} \\
&= - \left\{ \log \frac{\mu}{2-\mu} - \frac{\mu}{2-\mu} \log(2) + \left(\frac{\mu}{2-\mu} - 1 \right) E \log(y) \right\} \\
&= - \left\{ \log \frac{\mu}{2-\mu} - \frac{\mu}{2-\mu} \log(2) + \left(\frac{\mu}{2-\mu} - 1 \right) \right. \\
&\quad \left. \left[-\frac{(2-\mu)}{\mu} + \log(2) \right] \right\} \\
&= -\log \frac{\mu}{2-\mu} + \frac{\mu}{2-\mu} \log(2) - \left(\frac{\mu}{2-\mu} - 1 \right) \left[-\frac{(2-\mu)}{\mu} + \log(2) \right] \\
&= -\log \frac{\mu}{2-\mu} + \frac{\mu}{2-\mu} \log(2) - \left[\left(\frac{\mu}{2-\mu} - 1 \right) \left(-\frac{(2-\mu)}{\mu} \right) \right] \\
&\quad - \left[\left(\frac{\mu}{2-\mu} - 1 \right) \log(2) \right] \\
&= \log \frac{\mu}{2-\mu} + 1 - \frac{(2-\mu)}{\mu} + \log(2).
\end{aligned}$$

-

APÊNDICE B – VETOR *ESCORE* E MATRIZ DE INFORMAÇÃO DE FISHER

B.1 VETOR *ESCORE*

Considere uma amostra independente $Y_1, \dots, Y_n$ tal que $Y_i \sim \text{GOLL-}\mathfrak{T}(\beta)$. Da expressão (3.1) no Capítulo 3, o logaritmo da função de verossimilhança (função de log-verossimilhança) é dada por:

$$\begin{aligned} \ell(\beta) &= \log \prod_{i=1}^n f(y_i; \beta) \\ &= \sum_{i=1}^n \log \left\{ \frac{\mu_i(\beta)}{2 - \mu_i(\beta)} 2^{-\frac{\mu_i(\beta)}{2 - \mu_i(\beta)}} y_i^{\frac{\mu_i(\beta)}{2 - \mu_i(\beta)} - 1} \right\} \\ &= \sum_{i=1}^n \log \frac{\mu_i(\beta)}{2 - \mu_i(\beta)} - \log(2) \sum_{i=1}^n \frac{\mu_i(\beta)}{2 - \mu_i(\beta)} + \sum_{i=1}^n \left[\frac{\mu_i(\beta)}{2 - \mu_i(\beta) - 1} \right] \log y_i. \end{aligned}$$

A *escore* é a derivada da função de log-verossimilhança em relação aos parâmetros β . A *escore* para a observação i , denotada por $U_i(\beta)$, é dada por:

$$U_i(\beta) = \frac{\partial \ell_i(\beta)}{\partial \beta} = \frac{\partial \ell_i(\beta)}{\partial \mu_i(\beta)} \frac{\partial \mu_i(\beta)}{\partial \eta_i(\beta)} \frac{\partial \eta_i(\beta)}{\partial \beta},$$

em que

$$\frac{\partial \mu_i(\beta)}{\partial \eta_i} = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} = \frac{1}{g'(\eta_i)} g'(\eta_i) = \frac{\partial g(\mu_i)}{\partial \mu_i}.$$

A expressão para a derivada parcial da log-verossimilhança em relação a $\mu_i(\beta)$ pode ser simplificada e reescrita como:

$$\begin{aligned} \frac{\partial \ell_i(\beta)}{\partial \mu_i(\beta)} &= \frac{(2 - \mu_i(\beta))}{\mu_i(\beta)} \left[\frac{2 - \mu_i(\beta) + \mu_i(\beta)}{(2 - \mu_i(\beta))} \right] - \log(2) \left[\frac{2 - \mu_i(\beta) + \mu_i(\beta)}{(2 - \mu_i(\beta))^2} \right] \\ &\quad + \log y_i \left[\frac{2 - \mu_i(\beta) + \mu_i(\beta)}{(2 - \mu_i(\beta))^2} \right] \\ &= \frac{(2 - \mu_i(\beta))}{\mu_i(\beta)} \frac{2}{(2 - \mu_i(\beta))^2} - \log(2) \frac{2}{(2 - \mu_i(\beta))^2} + \log y_i \frac{2}{(2 - \mu_i(\beta))^2} \\ &= \frac{2}{(2 - \mu_i(\beta))^2} \left[\frac{(2 - \mu_i(\beta))}{\mu_i(\beta)} + \log y_i - \log(2) \right], \end{aligned}$$

considerando

$$\delta_i(\boldsymbol{\beta}) = \frac{2}{(2 - \mu_i(\boldsymbol{\beta}))^2}$$

e

$$p_i(\boldsymbol{\beta}) = -\frac{2 - \mu_i}{\mu_i} + \log 2,$$

a derivada da log-verossimilhança em relação a $\mu_i(\boldsymbol{\beta})$ pode ser expressa como:

$$\frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \mu_i(\boldsymbol{\beta})} = \delta_i [\log y_i - p_i(\boldsymbol{\beta})].$$

A derivada de $\eta_i(\boldsymbol{\beta})$ em relação a β_k é dada por:

$$\frac{\partial \eta_i(\boldsymbol{\beta})}{\partial \beta_k} = \mathbf{x}_{ik},$$

a função *escore* para i -ésima observação é expressa por

$$\mathbf{U}_i(\boldsymbol{\beta}) = \mathbf{x}_i \frac{1}{g'(\mu_i)} \delta_i(\boldsymbol{\beta}) [\log y_i - p_i(\boldsymbol{\beta})],$$

assim, a função *escore* na forma matricial para n observações é dada por:

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{U}_i(\boldsymbol{\beta}) = \mathbf{X}^\top \mathbf{W} \boldsymbol{\Delta} (\log \mathbf{y} - \mathbf{p}(\boldsymbol{\beta})),$$

em que $\mathbf{X} = [\mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_n]^\top$ é a matriz modelo $n \times (p + 1)$ tal que $\mathbf{x}_i$ representa o vetor de covariáveis para a i -ésima observação, $\mathbf{W} = \text{diag}\left(\frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)}\right)$, tal que $\text{diag}(\cdot)$ representa uma matriz diagonal, $\boldsymbol{\Delta} = \text{diag}(\delta_1(\boldsymbol{\beta}), \dots, \delta_n(\boldsymbol{\beta}))$, $\log \mathbf{y} \equiv [\log y_1, \dots, \log y_n]^\top$ e $\mathbf{p}(\boldsymbol{\beta}) = [p_1(\boldsymbol{\beta}), \dots, p_n(\boldsymbol{\beta})]^\top$.

B.1.1 Esperança do *escore*

A esperança do *escore* pode ser calculada como:

$$E[\mathbf{U}_i(\boldsymbol{\beta})] = E\left[\frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right].$$

A esperança da função *escore* é igual a zero, ou seja,

$$E[U_i(\beta)] = 0.$$

A *escore* $U_i(\beta)$ é dada por:

$$U_i(\beta) = \frac{\partial \ell_i(\beta)}{\partial \beta},$$

onde $\ell_i(\beta)$ é a log-verossimilhança da i -ésima observação.

Temos que, para um modelo de máxima verossimilhança

$$E\left[\frac{\partial \ell_i(\beta)}{\partial \beta}\right] = 0.$$

Para demonstrar os cálculos da *escore*, começamos com a seguinte equação:

$$U = S[\log y - p(\beta)].$$

Onde $\log y = (\log y_1, \dots, \log y_n)$, $p(\beta) = (p_1(\beta), \dots, p_n(\beta))$ e $S = X^\top W \Delta$, a matriz S é uma combinação das matrizes $X^\top$ é transposta da matriz das variáveis explicativas, W é matriz diagonal de pesos e Δ é matriz diagonal adicional relacionada ao modelo. Assim, temos que a esperança da *escore* é

$$EU = E[S(\log y - p(\beta))] = S[E(\log y) - p(\beta)],$$

em que, $E(\log y_i) = -\frac{2-\mu_i}{\mu_i} + \log 2$ e $p_i(\beta) = -\frac{2-\mu_i}{\mu_i} + \log 2$, para $i = 1, \dots, n$. Logo,

$$EU = S[E(\log y) - p(\beta)] = 0.$$

B.1.2 Derivada do vetor *escore* para obter a Informação de Fisher

O vetor *escore* U para um modelo com parâmetros β é dado por:

$$U = \mathbf{X}^\top \mathbf{W} \Delta (\log \mathbf{y} - \mathbf{p}(\boldsymbol{\beta}))$$

derivando o vetor *escore* em relação a $\boldsymbol{\beta}^\top = [\beta_0, \dots, \beta_p]^\top$, temos

$$\frac{\partial U}{\partial \boldsymbol{\beta}^\top} = \frac{\partial}{\partial \boldsymbol{\beta}^\top} [\mathbf{X}^\top \mathbf{W} \Delta (\log \mathbf{y} - \mathbf{p}(\boldsymbol{\beta}))].$$

Fazendo a derivada $\frac{\partial U}{\partial \boldsymbol{\beta}^\top}$, para obter a informação de Fisher

$$\frac{\partial U}{\partial \boldsymbol{\beta}^\top} = -\mathbf{X}^\top \mathbf{W} \Delta \frac{\partial \mathbf{p}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} + \mathbf{X}^\top \frac{\partial (\mathbf{W} \Delta)}{\partial \boldsymbol{\beta}^\top} (\log \mathbf{y} - \mathbf{p}(\boldsymbol{\beta})).$$

A derivada de $\mathbf{W}$ é dada em relação a μ_i , sabendo que $\mathbf{W} = \text{diag} \left(\frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)} \right)$ e $g(\mu_i) = \log \left(\frac{\mu_i}{2 - \mu_i} \right)$,

fazendo termo a termo para a i -ésima observação temos que

$$\begin{aligned} g'(\mu_i) &= \frac{1}{\left(\frac{\mu_i}{2 - \mu_i} \right)} \frac{1(2 - \mu_i) - \mu_i(-1)}{(2 - \mu_i)^2} \\ &= \frac{1}{\mu_i} (2 - \mu_i) \frac{2 - \mu_i + \mu_i}{(2 - \mu_i)^2} \\ &= \frac{2(2 - \mu_i)}{\mu_i(2 - \mu_i)^2} = \frac{2}{\mu_i(2 - \mu_i)}, \end{aligned}$$

assim, temos que,

$$\frac{1}{g'(\mu_i)} = \frac{1}{\frac{2}{\mu_i(2 - \mu_i)}} = \frac{\mu_i(2 - \mu_i)}{2}.$$

Fazendo a derivada de $\delta(\boldsymbol{\beta})$, temos:

$$\frac{\partial \delta}{\partial \boldsymbol{\beta}^\top} = \frac{\partial \delta}{\partial \boldsymbol{\mu}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}^\top},$$

começando com $\frac{\partial \delta}{\partial \boldsymbol{\mu}}$:

$$\begin{aligned}
\frac{\partial \delta}{\partial \mu} &= \left[\frac{2}{(2 - \mu_i(\beta))^2} \right]' \Leftrightarrow [2(2 - \mu_i(\beta))^{-2}]' \\
&= -2 \cdot 2(2 - \mu_i(\beta))^{-3} \cdot (-1) \\
&= -4(2 - \mu_i(\beta))^{-3} \cdot (-1) \\
&= 4(2 - \mu_i(\beta))^{-3} \\
&= \frac{4}{(2 - \mu_i(\beta))^3}.
\end{aligned}$$

Para encontrar o $\frac{\partial p(\beta)}{\partial \beta^\top}$, encontra-se à as seguintes derivadas

$$\frac{\partial p(\beta)}{\partial \beta^\top} = \frac{\partial p}{\partial \mu} \frac{\partial \mu}{\partial \eta} \frac{\partial \eta}{\partial \beta}.$$

fazendo a derivada de $\frac{\partial p(\beta)}{\partial \mu}$, temos que:

$$\begin{aligned}
\frac{\partial p(\beta)}{\partial \mu} &= \left[-\frac{(2 - \mu_i)}{\mu_i} + \log(2) \right]' \\
&= \frac{[-(-1)\mu_i] - [-(2 - \mu_i)(1)]}{\mu_i^2} \\
&= \frac{\mu_i + 2 - \mu_i}{\mu_i^2} \\
&= \frac{2}{\mu_i^2}.
\end{aligned}$$

A seguir, a derivada de $\frac{\partial \mu}{\partial \eta}$, é obtida a partir da função logística

$$\mu = \frac{2e^\eta}{1 + e^\eta},$$

derivando em η :

$$\begin{aligned}
\frac{\partial \mu}{\partial \eta} &= \left[\frac{2e^\eta}{(1 + e^\eta)} \right]' \\
&= \frac{2e^\eta(1 + e^\eta) - 2e^\eta e^\eta}{(1 + e^\eta)^2} \\
&= \frac{2e^\eta + 2e^{2\eta} - 2e^{2\eta}}{(1 + e^\eta)^2} \\
&= \frac{2e^\eta}{(1 + e^\eta)^2} \\
&= \frac{2e^\eta}{(1 + e^\eta)} \frac{1}{(1 + e^\eta)}.
\end{aligned}$$

sabe-se que $\mu = \frac{2e^n}{1+e^n}$, e $\frac{1}{(1+e^n)} = \frac{(2-\mu)}{2}$ logo, $\frac{\partial \mu}{\partial \eta} = \frac{\mu(2-\mu)}{2}$.
provando que $\frac{\partial \eta}{\partial \beta}$, em que $\eta_i = \beta_0 + \beta_1 X_{1i}, \dots, \beta_k X_{ki}$

$$\begin{aligned}\frac{\partial \eta}{\partial \beta} &= 1 + 0 \\ \frac{\partial \eta}{\partial \beta_1} &= 0 + X_{1i} \\ \frac{\partial \eta}{\partial \beta_2} &= 0 + 0 + X_{2i} \\ &\vdots \\ \frac{\partial \eta}{\partial \beta_{ki}} &= 0 + 0 + \dots + X_{ki},\end{aligned}$$

logo, $\frac{\partial \eta}{\partial \beta} = \mathbf{X}$.

Temos

$$\begin{aligned}\frac{\partial \mathbf{p}_i(\beta)}{\partial \beta^\top} &= -\frac{1}{2} \frac{\partial}{\partial \beta^\top} \left\{ \frac{[2 - \mu_i(\beta)]^2}{\mu_i(\beta)} \right\} - \frac{1}{2} \frac{\partial [2 - \mu_i(\beta)]}{\partial \beta^\top} \\ &= -\frac{1}{2} \left\{ \frac{-2 \frac{\partial \mu_i(\beta)}{\partial \beta^\top} [2 - \mu_i(\beta)] - \frac{\partial \mu_i(\beta)}{\partial \beta^\top} [2 - \mu_i(\beta)]^2}{\mu_i(\beta)^2} \right\} + \frac{1}{2} \frac{\partial \mu_i(\beta)}{\partial \beta^\top} \\ &= \frac{1}{2} \frac{\partial \mu_i(\beta)}{\partial \beta^\top} + \frac{1}{2} \frac{[2 - \mu_i(\beta)]}{\mu_i^2(\beta)} \frac{\partial \mu_i(\beta)}{\partial \beta^\top} [1 + 2 - \mu_i(\beta)].\end{aligned}$$

Podemos expressar a matriz de informação de Fisher como:

$$\mathbf{K}(\beta) = \mathbf{X}^\top \mathbf{W} \mathbf{\Delta} \mathbf{V},$$

em que $\mathbf{V} = \frac{\partial \mathbf{p}(\beta)}{\partial \beta^\top}$ é uma matriz $n \times (p+1)$, dada por:

$$\mathbf{V} = \begin{bmatrix} \frac{2}{\mu_1^2} \frac{2-\mu_1}{\mu_1} \mathbf{x}_1^\top \\ \frac{2}{\mu_2^2} \frac{2-\mu_2}{\mu_2} \mathbf{x}_2^\top \\ \vdots \\ \frac{2}{\mu_n^2} \frac{2-\mu_n}{\mu_n} \mathbf{x}_n^\top \end{bmatrix}.$$