



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE TECNOLOGIA**  
**DEPARTAMENTO DE ENGENHARIA ELÉTRICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**  
**MESTRADO ACADÊMICO EM ENGENHARIA ELÉTRICA**

**ANA CAROLINA ARAÚJO DE OLIVEIRA**

**PROPOSTA DE UM MODELO PREDITIVO DE RESULTADOS DE LEILÕES DE  
TRANSMISSÃO DE ENERGIA ELÉTRICA**

**FORTALEZA**

**2023**

ANA CAROLINA ARAÚJO DE OLIVEIRA

PROPOSTA DE UM MODELO PREDITIVO DE RESULTADOS DE LEILÕES DE  
TRANSMISSÃO DE ENERGIA ELÉTRICA

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia Elétrica do Programa de Pós-Graduação em Engenharia Elétrica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Elétrica. Área de Concentração: Engenharia Elétrica.

Orientador: Prof. Dr. Ricardo Silva Thé Pontes.

Coorientador: Dr. Francisco Alexandre Andrade de Souza.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- O45p Oliveira, Ana Carolina Araújo de.  
Proposta de um modelo preditivo de resultados de leilões de transmissão de energia elétrica / Ana Carolina Araújo de Oliveira. – 2023.  
80 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Fortaleza, 2023.  
Orientação: Prof. Dr. Ricardo Silva Thé Pontes.
1. RAP (Receita Anual Permitida). 2. Leilão de Transmissão de Energia. 3. Teoria dos Leilões. 4. Regressão Linear Bayesiana. 5. Deságio. I. Título.

CDD 621.3

---

A Deus e à minha família.

## AGRADECIMENTOS

À Deus, por ter me dado saúde e força permitindo a conclusão desta pesquisa.

À minha mãe, Lourdes, pelos cuidados e amor incondicional.

Ao meu pai, Agenor, pelo bom humor nas horas difíceis.

À minha irmã, Fernanda, pelo seu apoio em todos os momentos e por ser minha referência de vida.

Ao meu irmão, Moisés, pelos cuidados e amor paternais.

Aos queridos Mateus Andrade e família, pela paciência e por tanto carinho.

Ao Departamento de Engenharia Elétrica por toda contribuição em meu crescimento.

Ao Prof. Dr. Ricardo Silva Thé Pontes, pela sua excelente orientação, conselhos e sua exímia dedicação ao ensino.

Ao Dr. Francisco Alexandre Andrade de Souza, pelos seus ensinamentos e por sua solicitude nos momentos em dúvida.

Ao Prof. Dr. Victor de Paula Brandão Aguiar pelos conselhos e pelo incentivo ao meu crescimento.

Aos colegas do Laboratório de Sistemas Motrizes (LAMOTRIZ) por tornarem esta caminhada mais leve e rica de aprendizados.

Aos participantes da banca examinadora Profa. Dra. Ruth Pastôra Saraiva Leão e Prof. Dr. Mauricio Benegas pelo tempo, pelas colaborações e sugestões.

À CAPES, pelo apoio financeiro com a manutenção da bolsa de auxílio.

Aos colegas da turma de mestrado, pelas reflexões, críticas e sugestões recebidas.

"As questões mais importantes da vida de fato, são, na maior parte, apenas problemas de probabilidade" (LAPLACE, P. S., séc. XIX.)

## RESUMO

A dissertação apresenta uma proposta de um modelo preditivo baseado em regressão linear bayesiana para analisar e prever os resultados da diferença entre a Receita Anual Permitida definida em edital pela Receita Anual Permitida proposta em leilões de transmissão de energia elétrica no Brasil. Em um cenário de crescimento acelerado na geração e no sistema de transmissão de energia, os leilões de transmissão desempenham um papel importante na expansão e modernização da infraestrutura elétrica nacional. O estudo destaca a importância estratégica desses leilões como um mecanismo fundamental para a alocação eficiente de recursos, garantindo a expansão da rede de transmissão de forma econômica e sustentável. Para este estudo, inicialmente foi utilizado um modelo de regressão linear múltiplo, onde obteve-se que as variáveis utilizadas neste modelo explicavam aproximadamente 46,8% da variabilidade da variável dependente. Após obter dados pouco satisfatórios, mudou-se para uma técnica de regressão linear bayesiana. Ao analisar dados históricos dos leilões de transmissão, o modelo bayesiano foi capaz de explicar aproximadamente 87% da variabilidade no valor do lance, que significa um valor alto quando comparado aos demais estudos e sabendo que a maioria das variáveis levantadas estão contidas nos editais dos leilões. É importante ressaltar que as variáveis mais relevantes identificadas pelo modelo foram as variáveis número de empreendimentos, extensão da linha de transmissão, empreendimentos realizados no subsistema nordeste, prazo para execução do empreendimento e número de competidores, indicando sua forte influência nos resultados dos leilões. Esta descoberta oferece percepções para os participantes do mercado, permitindo-lhes entender melhor quais fatores podem impactar suas estratégias de licitação. A relevância deste estudo não se limita apenas à sua aplicação prática. Ele destaca a necessidade crucial de uma compreensão aprofundada dos leilões de transmissão no contexto do desenvolvimento energético brasileiro. Ao estudar esses eventos, não apenas como transações econômicas, mas como indicadores sensíveis das dinâmicas do mercado, pode-se promover um setor de energia mais eficiente e competitivo no Brasil. Esta dissertação não apenas propõe um modelo preditivo usando regressão linear bayesiana para os leilões de transmissão de energia elétrica no Brasil, mas também ressalva a importância estratégica de compreender esses leilões, contribuindo para um conhecimento mais amplo sobre os mecanismos do mercado de energia elétrica.

**Palavras-chave:** RAP (Receita Anual Permitida); Leilão de Transmissão de Energia; Teoria dos Leilões; Regressão Linear Bayesiana; Deságio.

## ABSTRACT

The dissertation presents a proposal for a predictive model based on Bayesian linear regression to analyze and predict the results of the difference between the Permitted Annual Revenue defined in the notice and the Permitted Annual Revenue proposed in electricity transmission auctions in Brazil. In a scenario of accelerated growth in energy generation and transmission system, transmission auctions play an important role in the expansion and modernization of the national electrical infrastructure. The study highlights the strategic importance of these auctions as a fundamental mechanism for the efficient allocation of resources, ensuring the expansion of the transmission network in an economic and sustainable way. For this study, a multiple linear regression model was initially used, where it was found that the variables used in this model explained approximately 46.8% of the variability of the dependent variable. After obtaining unsatisfactory data, he switched to a Bayesian linear regression technique. When analyzing historical data from transmission auctions, the Bayesian model was able to explain approximately 87% of the variability in the bid value, which means a high value when compared to other studies and knowing that most of the variables raised are contained in the auction notices. It is important to highlight that the most relevant variables identified by the model were the variables number of projects, length of the transmission line, projects carried out in the northeast subsystem, deadline for executing the project and number of competitors, indicating their strong influence on the results of the auctions. This finding offers insights to market participants, allowing them to better understand which factors may impact their bidding strategies. The relevance of this study is not limited to its practical application. It highlights the crucial need for an in-depth understanding of transmission auctions in the context of Brazilian energy development. By studying these events, not just as economic transactions, but as sensitive indicators of market dynamics, a more efficient and competitive energy sector in Brazil can be promoted. This dissertation not only proposes a predictive model using Bayesian linear regression for electricity transmission auctions in Brazil, but also highlights the strategic importance of understanding these auctions, contributing to broader knowledge about the mechanisms of the electricity market.

**Keywords:** RAP (Allowed Annual Revenue); Energy Transmission Auction; Auction Theory; Bayesian Linear Regression; Discount.



## LISTA DE FIGURAS

Figura 1 – Oferta interna de energia elétrica por fonte - 2009 . . . . .	14
Figura 2 – Oferta interna de energia elétrica por fonte - 2022 . . . . .	14
Figura 3 – Expansão da extensão das linhas de transmissão no Brasil - 1999 a 2022 . .	15
Figura 4 – Investimentos em transmissão de energia - PDE 2032 . . . . .	17
Figura 5 – Investimentos por LT e SE . . . . .	18
Figura 6 – Estimativa de Expansão de consumo de eletricidade na rede . . . . .	20
Figura 7 – Diagrama de árvore - Dilema dos Prisioneiros . . . . .	31
Figura 8 – Edital N° 1/2023-ANEEL . . . . .	36
Figura 9 – Sumário do Edital N° 1/2023-ANEEL . . . . .	37
Figura 10 – Diagrama causal - Preço do Lote . . . . .	43
Figura 11 – Mapa de calor da correlação entre as variáveis . . . . .	61
Figura 12 – Código no Matlab utilizado para modelagem . . . . .	72

## LISTA DE TABELAS

Tabela 1 – Estimativa de Expansão das LTs - PDE 2032 . . . . .	19
Tabela 2 – Estimativa de Expansão das SEs - PDE 2032 . . . . .	19
Tabela 3 – Matriz de <i>payoffs</i> . . . . .	29
Tabela 4 – Exemplo de matriz de <i>payoffs</i> . . . . .	30
Tabela 5 – Média e desvio padrão . . . . .	60
Tabela 6 – Correlação entre as variáveis . . . . .	61
Tabela 7 – Resultados obtidos - Regressão Linear Múltipla . . . . .	61
Tabela 8 – Resultados obtidos - Regressão Linear Bayesiana . . . . .	63
Tabela 9 – Saída do modelo no Stan . . . . .	64

## **LISTA DE ABREVIATURAS E SIGLAS**

ANEEL	Agência Nacional de Energia Elétrica
CNAEE	Conselho Nacional de Águas e Energia Elétrica
EPE	Empresa de Pesquisa Energética
MME	Ministério de Minas e Energia
ONS	Operador Nacional do Sistema Elétrico
RAP	Receita Anual Permitida
SIN	Sistema Interligado Nacional
SPE	Secretaria de Planejamento e Desenvolvimento Energético
SPG	Secretaria de Petróleo, Gás Natural e Biocombustíveis
SSE	Soma dos Quadrados dos Resíduos
SSR	Soma dos Quadrados de Regressão
SST	Soma Total os Quadrados

## LISTA DE SÍMBOLOS

<i>MVA</i>	Potência
<i>km</i>	Extensão
<i>kV</i>	Tensão

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
<b>1.1</b>	<b>Contextualização</b>	<b>13</b>
<i>1.1.1</i>	<i>Transição energética</i>	<i>13</i>
<i>1.1.2</i>	<i>Sistema de transmissão nacional</i>	<i>15</i>
<i>1.1.3</i>	<i>Planejamento da transmissão</i>	<i>16</i>
<i>1.1.4</i>	<i>Leilões de transmissão - breve histórico</i>	<i>20</i>
<b>1.2</b>	<b>Motivação</b>	<b>21</b>
<b>1.3</b>	<b>Objetivo geral</b>	<b>22</b>
<b>1.4</b>	<b>Objetivos específicos</b>	<b>22</b>
<b>1.5</b>	<b>Conteúdo da dissertação</b>	<b>23</b>
<b>2</b>	<b>TEORIA DOS LEILÕES</b>	<b>24</b>
<b>2.1</b>	<b>Introdução à Teoria dos Jogos</b>	<b>24</b>
<i>2.1.1</i>	<i>Histórico</i>	<i>24</i>
<i>2.1.2</i>	<i>Elementos da teoria dos jogos</i>	<i>26</i>
<i>2.1.3</i>	<i>Regras de um jogo</i>	<i>27</i>
<b>2.2</b>	<b>Teoria dos Leilões como uma parte da Teoria dos Jogos</b>	<b>31</b>
<b>2.3</b>	<b>Técnicas de Leilões</b>	<b>33</b>
<b>2.4</b>	<b>Leilões de Transmissão no Brasil</b>	<b>34</b>
<b>2.5</b>	<b>Editais do Leilão</b>	<b>36</b>
<b>3</b>	<b>ANÁLISE DOS LEILÕES DE TRANSMISSÃO</b>	<b>39</b>
<b>3.1</b>	<b>Revisão Bibliográfica</b>	<b>39</b>
<b>3.2</b>	<b>Variáveis do Modelo</b>	<b>41</b>
<b>3.3</b>	<b>Análise de modelos estatísticos</b>	<b>45</b>
<i>3.3.1</i>	<i>Modelo de Regressão Linear Múltipla</i>	<i>45</i>
<i>3.3.2</i>	<i>Modelo de Regressão Linear Bayesiana</i>	<i>54</i>
<b>4</b>	<b>RESULTADOS</b>	<b>60</b>
<b>4.1</b>	<b>Resultados do Modelo de Regressão Linear Múltiplo</b>	<b>60</b>
<b>4.2</b>	<b>Resultados do Modelo de Regressão Linear Bayesiana</b>	<b>62</b>
<b>5</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS</b>	<b>66</b>
<b>5.1</b>	<b>Considerações Finais</b>	<b>66</b>

<b>5.2</b>	<b>Trabalhos Futuros</b> . . . . .	<b>67</b>
<b>5.3</b>	<b>Publicações</b> . . . . .	<b>67</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>69</b>
	<b>APÊNDICE A – MODELAGEM - REGRESSÃO LINEAR MÚLTIPLA</b>	<b>72</b>
	<b>APÊNDICE B – MODELAGEM - REGRESSÃO LINEAR BAYESIANA</b>	<b>73</b>

# 1 INTRODUÇÃO

Neste capítulo serão apresentados uma contextualização sobre o tema de pesquisa, a motivação, os objetivos geral e específicos e o conteúdo da dissertação.

## 1.1 Contextualização

Antes de explanar sobre os leilões de energia e suas características, é de fundamental importância que seja feita uma contextualização sobre energia no Brasil, transição energética, sistema de transmissão e planejamento de transmissão. Estes assuntos serão tratados neste tópico.

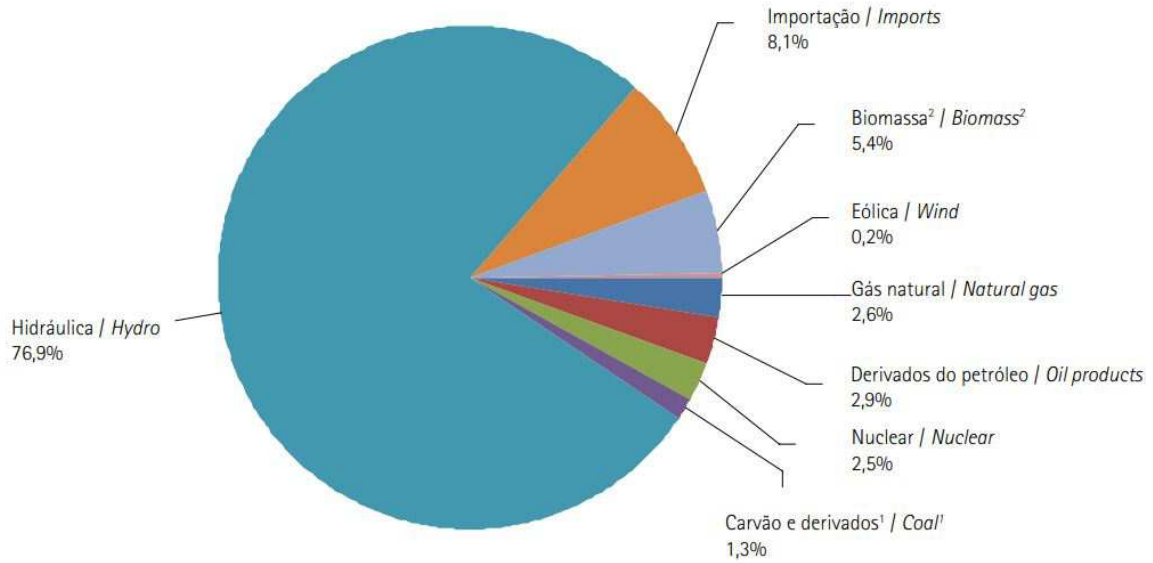
### 1.1.1 *Transição energética*

No últimos anos o Brasil vem passando por uma transformação significativa em sua matriz energética e infraestrutura de transmissão para atender às crescentes demandas por energia elétrica.

Tradicionalmente a matriz da energia elétrica brasileira é predominantemente hidrelétrica, entretanto, o país vem diversificando suas fontes de energia para garantir sustentabilidade e estabilidade no fornecimento. As Figuras 1 e 2 que correspondem, respectivamente à oferta interna de energia elétrica por fonte nos anos de 2009 e 2022, analisando as duas figuras, pode-se perceber que no período de 13 anos a matriz energética brasileira continua sendo majoritariamente hidráulica, como era de se esperar devido à disponibilidade desta fonte, porém, é possível notar que a participação de fonte eólica aumentou de 0,2 para 11,8%, um crescimento de 59 vezes comparado ao primeiro período. Além disso, em 2009 não havia participação da fonte solar, que em 2022 já correspondia à 4,4% da geração de energia elétrica.

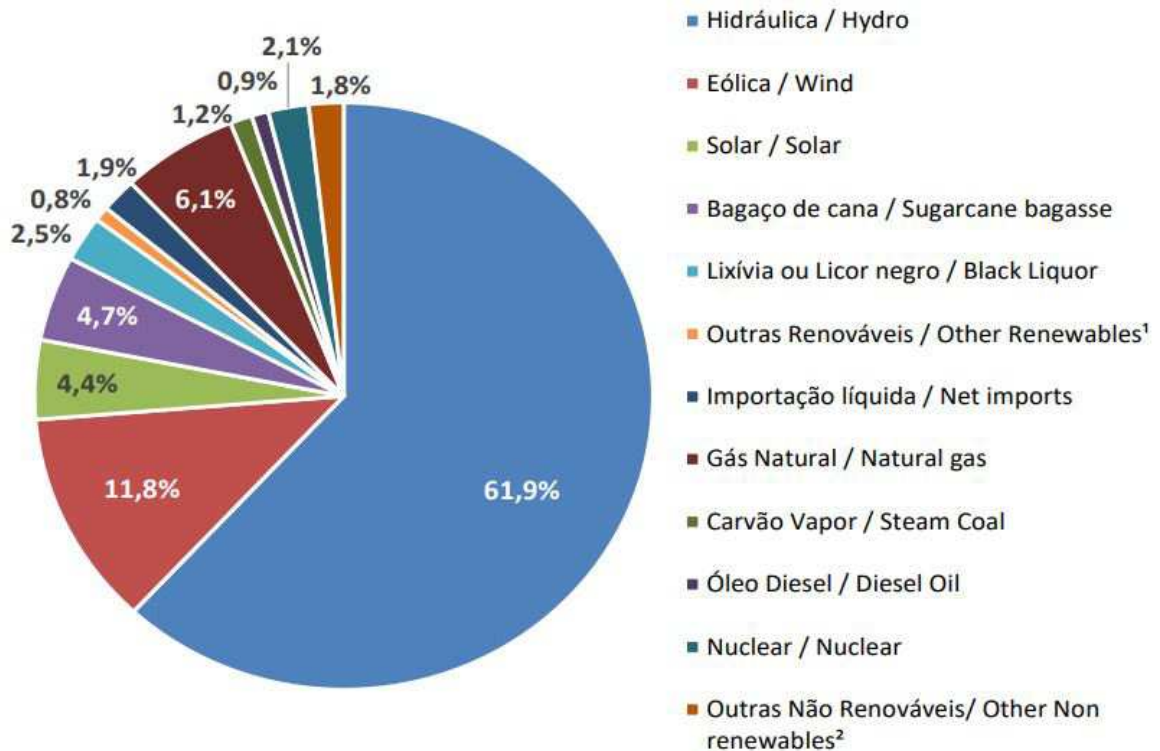
A expansão da matriz energética envolveu investimentos em fontes de energia renovável, como eólica e solar. Parques eólicos, especialmente no Nordeste, capitalizaram os ventos constantes da região, enquanto projetos solares foram implementados em várias partes do país, aproveitando a intensa incidência solar. Além disso, iniciativas envolvendo biomassa, como bagaço de cana-de-açúcar, e biogás a partir de resíduos orgânicos também foram aproveitadas, contribuindo para uma geração de energia mais sustentável. Essa mudança foi impulsionada por políticas públicas de incentivo à produção e consumo de energias renováveis. Leilões de energia e incentivos fiscais foram implementados para promover a participação do setor privado e acelerar o desenvolvimento de projetos sustentáveis.

Figura 1 – Oferta interna de energia elétrica por fonte - 2009



Fonte: (BEN, 2010)

Figura 2 – Oferta interna de energia elétrica por fonte - 2022



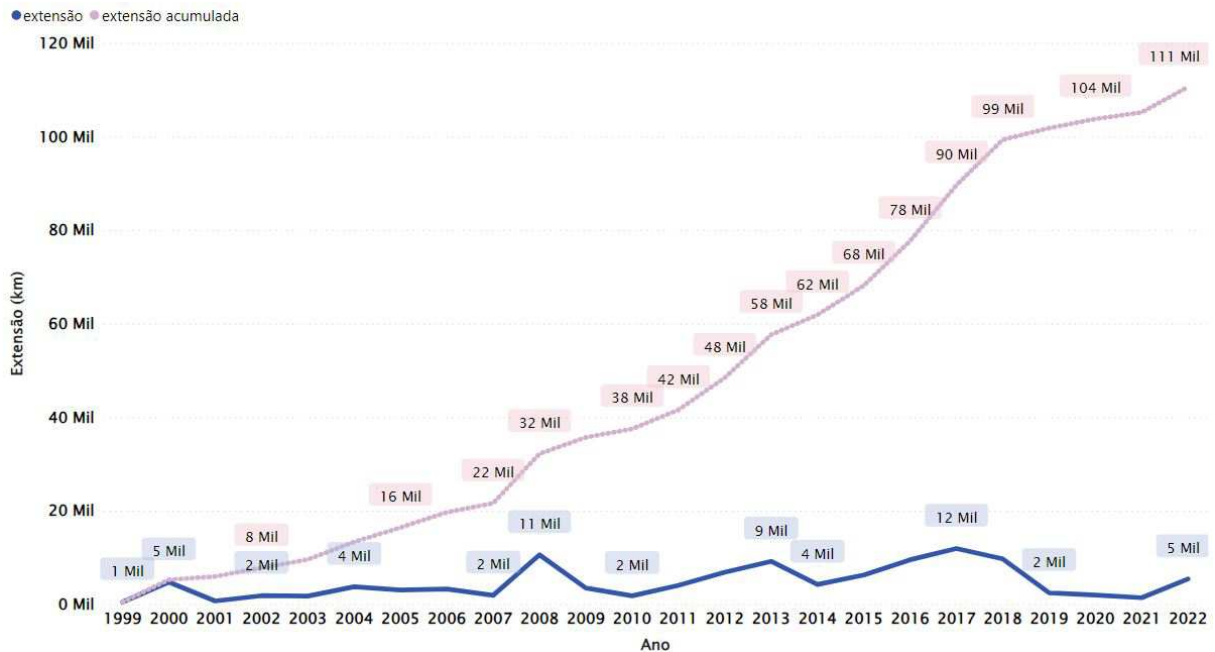
Fonte: (BEN, 2023)

Paralelamente à expansão da matriz energética, o Brasil investiu massivamente na expansão das linhas de transmissão. Com a necessidade de transportar eletricidade das novas fontes de geração até os centros de consumo, linhas de transmissão de alta tensão e subestações foram construídas em todo o país. Isso não apenas garantiu a distribuição eficiente de energia,



mas também permitiu a integração regional, possibilitando o compartilhamento de energia entre diferentes estados e regiões do Brasil. Como pode-se perceber na Figura 3, a extensão de linhas de transmissão construídas por ano entre os anos de 1999 e 2022, representada pela linha contínua azul, e a extensão acumulada, representada pela linha rosa pontilhada, totalizando 111 mil km de linhas de extensão construídos ao longo de 23 anos.

Figura 3 – Expansão da extensão das linhas de transmissão no Brasil - 1999 a 2022



Fonte: Elaborado pelo autor.

### 1.1.2 Sistema de transmissão nacional

O Brasil possui um sistema de transmissão nacional complexo e extenso, que inclui tanto sistemas interligados quanto sistemas isolados para atender às diferentes regiões do país. De acordo com (ONS, 2023a), o Sistema Interligado Nacional (SIN) é um sistema hidro-térmico-eólico de grande porte de produção e transmissão de energia elétrica, representando a maior parte da infraestrutura de transmissão sendo responsável por conectar as principais áreas de geração e consumo. Por outro lado, os Sistemas Isolados atendem áreas remotas ou de difícil acesso, onde a integração ao SIN é inviável ou economicamente inviável.

O SIN é o sistema de transmissão que conecta a maior parte do país. Ele é gerido pelo Operador Nacional do Sistema Elétrico (ONS) e integra diversas usinas de geração, subestações, linhas de transmissão e centros de consumo em uma única rede. A interconexão do SIN permite a transferência de energia elétrica entre diferentes regiões, garantindo o abastecimento

em áreas de alta demanda e permitindo a redistribuição de energia conforme as necessidades sazonais e emergenciais. A capacidade instalada de geração do sistema interligado é composta, principalmente, por usinas hidrelétricas distribuídas em dezesseis bacias hidrográficas nas diferentes regiões do país (ONS, 2023b). A instalação de usinas eólicas, principalmente nas regiões Nordeste e Sul, vem apresentando significativo crescimento, aumentando a importância dessa geração para o atendimento do mercado. Já as usinas térmicas geralmente são localizadas próximas aos centros de carga, garantindo segurança ao SIN quando se faz necessário que elas sejam despachadas, permitindo, assim, que se possa realizar uma melhor gestão dos estoques de água armazenada nos reservatórios das usinas hidrelétricas, garantindo um atendimento futuro da demanda.

Os Sistemas Isolados, por outro lado, são usados para atender áreas geograficamente distantes ou isoladas do sistema interligado. Isso inclui comunidades rurais remotas, pequenas cidades na região amazônica e outras áreas de difícil acesso. Esses sistemas muitas vezes dependem de fontes de energia locais, como usinas a diesel, energia solar ou eólica, devido à inviabilidade econômica ou logística de conectá-los ao SIN. Atualmente, no Brasil existem 212 localidades isoladas, sendo a maior parte está na região Norte, nos estados de Rondônia, Acre, Amazonas, Roraima, Amapá e Pará. Além da região norte tem também a ilha de Fernando de Noronha, em Pernambuco, e algumas localidades de Mato Grosso. Já em relação às capitais, Boa Vista (RR) é a única que ainda é atendida por um sistema isolado. O consumo nessas localidades é baixo e representa menos de 1% da carga total do país. A demanda por energia dessas regiões é suprida, principalmente, por térmicas a óleo diesel (ONS, 2023c).

A expansão e manutenção do sistema de transmissão nacional são essenciais para garantir um fornecimento de energia confiável em todo o país. No entanto, o Brasil enfrenta desafios como a necessidade de modernizar a infraestrutura existente, melhorar a eficiência da transmissão, enfrentar questões ambientais relacionadas à construção de novas linhas e subestações, e expandir a rede para áreas rurais e remotas.

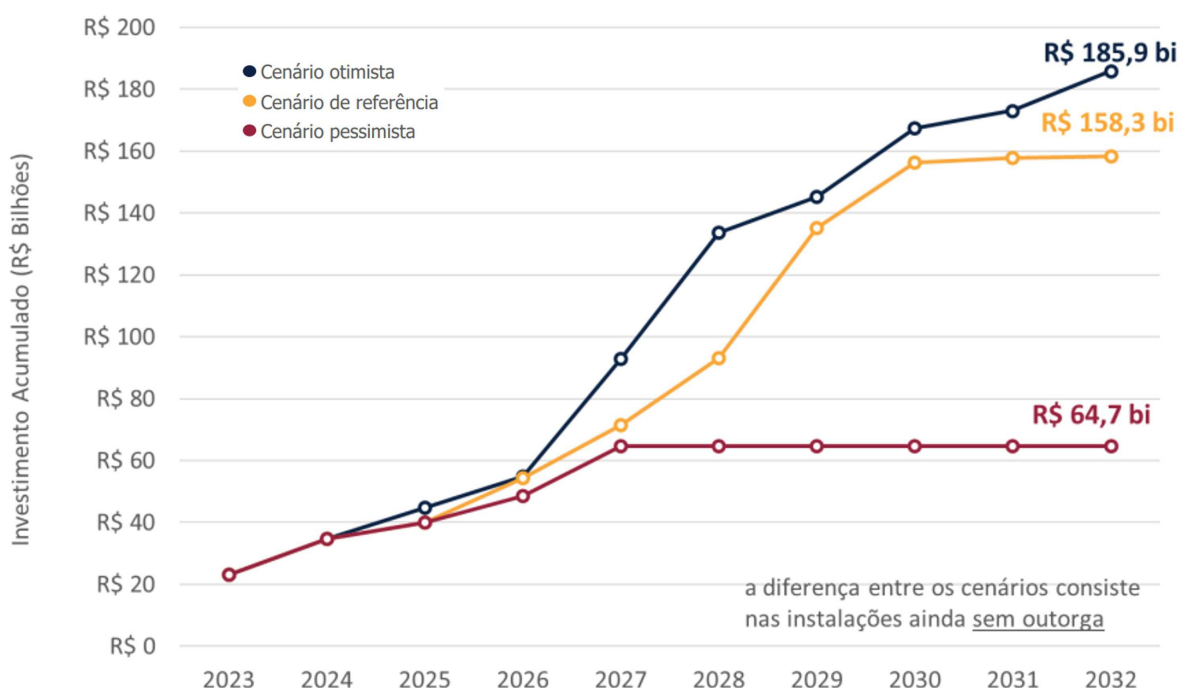
### ***1.1.3 Planejamento da transmissão***

O Plano Decenal de Energia (PDE) 2032 (EPE, 2023a), publicado em 2023 consiste em um estudo elaborado anualmente pela Empresa de Pesquisa Energética (EPE), sob as diretrizes e o apoio das equipes do Ministério de Minas e Energia (MME), sendo coordenados pelas Secretaria de Planejamento e Desenvolvimento Energético (SPE)/MME e de Secretaria de

Petróleo, Gás Natural e Biocombustíveis (SPG)/MME. Estes estudos funcionam como base para decisões de política energética e fornecem ao mercado as informações que possibilitem a análise do desenvolvimento do sistema elétrico e condições de suprimento para diferentes possíveis cenários.

De acordo com o PDE 2032, em um cenário otimista, espera-se que seja realizado um investimento total de R\$ 185,9 bi no período de 2023 a 2032, total resultante da implantação de obras ainda sem outorga, sendo considerada a data de necessidade inicialmente prevista nos estudos de planejamento. Vale ressaltar que as obras com outorga são aquelas que já têm permissão para construção, seja esta estadual, municipal, federal entre outras; já as obras sem outorga ainda estão aguardando esta licença, dessa forma, é essencial que estas licenças sejam aprovadas para a segurança no sistema de transmissão. Em um cenário de referência, é reavaliada a data de necessidade das obras sem outorgas, sendo feito uma atualização dessas datas, dessa forma, nesta circunstância, os investimentos previstos são de R\$ 158,3 bi em dez anos. De forma geral, o cenário referência considera cerca de 77% das obras sem outorgas que foram representadas no cenário otimista (EPE, 2023a). Já no cenário pessimista, esses investimentos entre 2023 e 2032 caíram para R\$64,7 bi, considerando que não haveria implantação das obras que ainda estão sem outorga, ou seja, a diferença entre os cenários consiste nas obras que ainda não têm outorga. Estes três cenários podem ser observados na Figura 4.

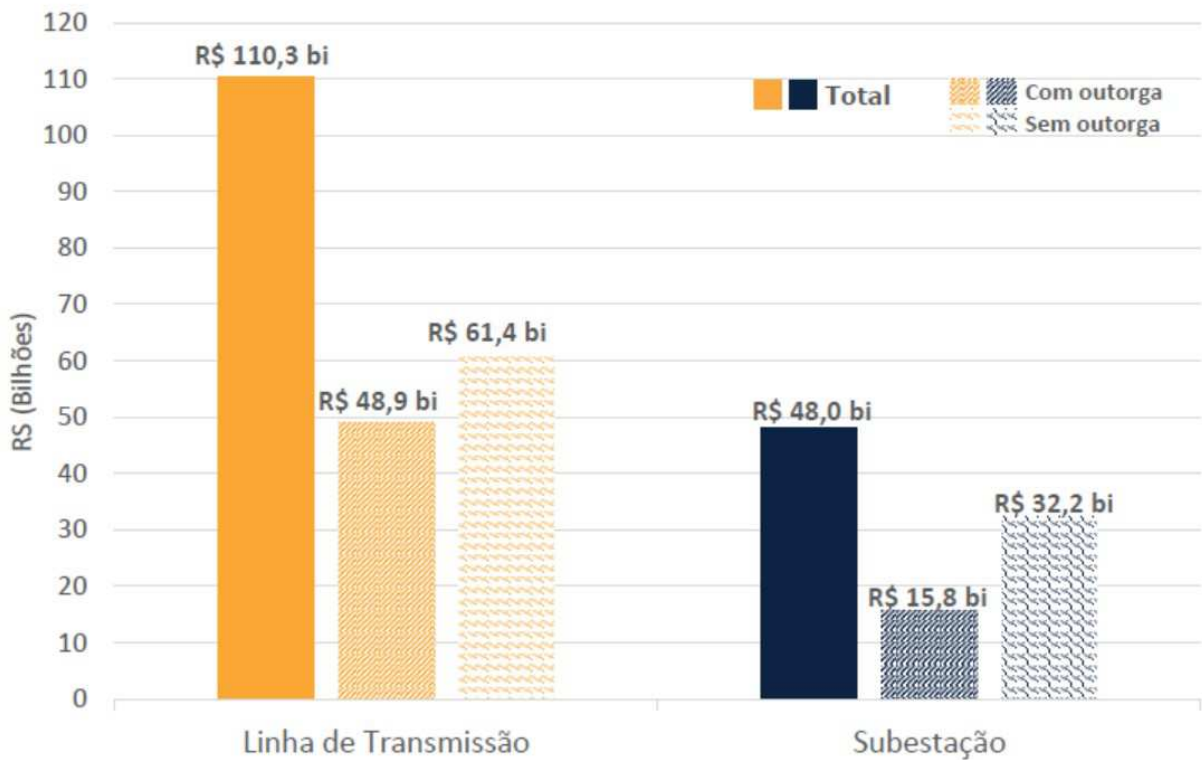
Figura 4 – Investimentos em transmissão de energia - PDE 2032



Fonte: Adaptado de (EPE, 2023a)

Como pode-se observar na Figura 5, para o período entre 2023 e 2032, de acordo com o cenário realista dos investimentos com outorga, serão destinados R\$ 48,9 bi para construção de linhas de transmissão e R\$15,8 bi para subestações. Já para as obras sem outorga, os investimentos serão de R\$ 61,4 bi para construção de linhas de transmissão e R\$ 32,2 bi para subestações. Dessa forma, no total, serão feitos investimentos de R\$ 158,3 bi em um cenário realista, onde R\$ 64,7 bi são para para obras com outorga e R\$ 93,6 bi para obras que ainda estão sem outorga. Portanto, pode-se observar uma grande diferença na previsão de investimentos nas instalações com e sem outorga.

Figura 5 – Investimentos por LT e SE



Fonte: (EPE, 2023a)

Quanto à previsão de extensão, ainda com base nos valores de referência, estima-se que até 2022 a extensão das linhas de transmissão totalizavam 179.297 km, entre os anos de 2023 e 2027 a previsão é de que a rede aumente 17.786 km e entre os anos de 2028 a 2032 aumente mais 23.674 km, totalizando uma extensão de 220.757 km ao final do período, o que representa um aumento de 41.460 km em 10 anos, correspondente a um aumento de 23% em relação à 2023. As estimativas de extensão das linhas de transmissão por faixa de tensão podem ser observadas na Tabela 1.

Em relação à previsão de expansão de construção das subestações, também com base

Tabela 1 – Estimativa de Expansão das LTs - PDE 2032

-	±800 kV	750 kV	±600 kV	500 kV	440 kV	345 kV	230 kV	Total
Estimativa dez/2021 (km)	9.204	2.683	12.816	70.031	6.859	11.092	66.612	179.297
Evolução 2023 - 2032 (km)	2.880	0	0	28.083	77	1.105	9.314	41.459
2023 - 2027 (km)	0	0	0	11.043	77	1.092	5.574	17.786
2028 - 2032 (km)	2.880	0	0	17.040	0	13	3.740	23.674
Estimativa dez/2032 (km)	12.084	2.683	12.816	98.114	6.936	12.198	75.926	220.757

Fonte: Adaptado de (EPE, 2023a).

nos valores de referência, estima-se que até 2022 a capacidade total instalada (estes valores são associados à capacidade instalada nominal dos transformadores referenciados ao lado de alta dos equipamentos) era de 461.286 MVA. Entre os anos de 2023 e 2027 a previsão é de que haja um aumento de 74.097 MVA e que entre os anos de 2028 a 2032 o acréscimo seja de 45.930 MVA, totalizando uma evolução de 581.313 MVA ao final do período, o que representa uma evolução de 120.027 MVA em 10 anos, correspondente à uma aumento de 26% em relação à 2023. As estimativas de expansão de capacidade instalada por faixa de tensão podem ser observadas na Tabela 2.

Tabela 2 – Estimativa de Expansão das SEs - PDE 2032

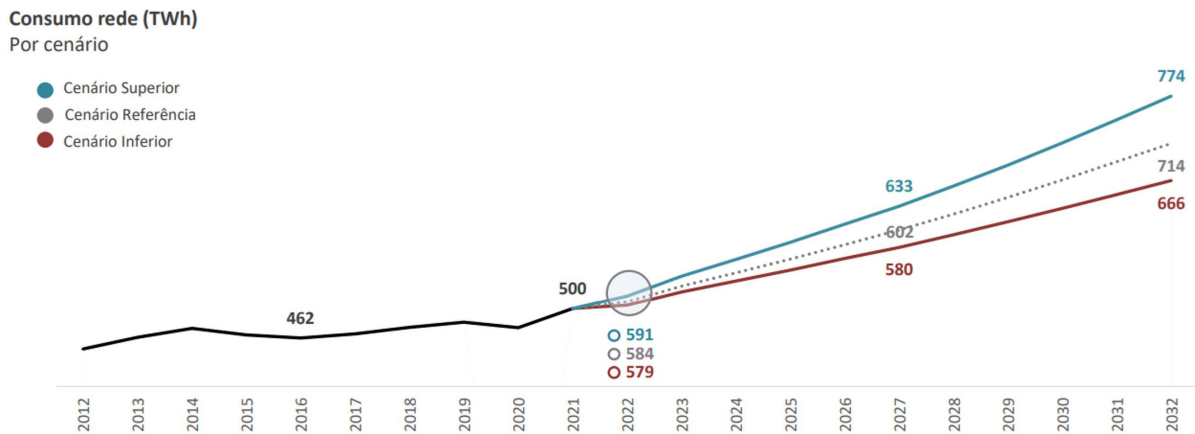
-	800 kV	750 kV	500 kV	440 kV	345 kV	230 kV	Total
Estimativa dez/2022 (MVA)	15.700	24.897	212.703	30.892	58.645	118.449	461.286
Evolução 2023 - 2032 (MVA)	10.000	0	63.758	1.681	14.602	29.986	120.027
2023 - 2027 (MVA)	0	0	38.336	1.231	13.803	20.727	74.097
2028 - 2032 (MVA)	10.000	0	25.422	450	799	9.259	45.930
Estimativa dez/2032 (MVA)	25.700	24.897	276.461	32.573	73.247	148.435	581.313

Fonte: Adaptado de (EPE, 2023a)

Todo esse aumento expressivo no sistema de transmissão de energia elétrica no Brasil é justificado pela estimativa de consumo. Conforme pode ser observado na Figura 6, foram estimados três cenários de consumo total de eletricidade. De acordo com o Plano Decenal de Expansão de Energia 2032 - Demanda de Eletricidade (EPE, 2023b), em um cenário superior, estima-se o consumo de 591 TWh ao final de 2022 e de 774 TWh em 2032, correspondendo à um aumento médio de 3,09% ao ano neste período. No cenário inferior, o consumo sairia de 579 TWh em 2022 para 666 TWh em 2032, o que representa um aumento de 1,50% de consumo. Por fim, no cenário referência, em 2022 a estimativa foi de 584 TWh e para 2032 é de 714 TWh, representando um crescimento de 130 TWh, o que equivale à média de 2,53% por ano.

A perspectiva de crescimento para a indústria é que, para um cenário de referência levando em consideração um consumo de 184 TWh em 2022 chegue à 243 TWh em 2032,

Figura 6 – Estimativa de Expansão de consumo de eletricidade na rede



Fonte: (EPE, 2023b)

representando um aumento de 59 TWh, ou em média 3,2% anuais. Neste cenário supõe-se o estímulo a investimentos em infraestrutura e perspectivas de bom desempenho para os setores produtores de *commodities*, além de ganhos de competitividades decorrentes da realização de reformas estruturais.

Nas grandes indústrias, como de cimentos, química (petroquímica e soda-cloro), papel e celulose (papel, celulose e pasta de alto rendimento) e metalurgia (alumínio primário, alumina, bauxita, aço bruto, pelotização, ferro-ligas e cobre), o consumo estimado aumenta de 64 TWh em 2022 para 81 TWh em 2032, representando em média 2,65% ao ano. Já no consumo comercial, em um cenário de referência, o consumo sai de 89 TWh em 2022 para 135 TWh em 2032, representando um aumento de 46 TWh em 10 anos e uma média de 5,17% ao ano.

#### 1.1.4 Leilões de transmissão - breve histórico

Os leilões de transmissão no Brasil fazem parte de uma estratégia governamental para promover o desenvolvimento da infraestrutura elétrica do país. No final da década de 1990, o Brasil adotou um modelo mais liberalizado para o setor energético, incentivando a participação do setor privado. Isso incluiu a implementação de leilões para concessões de projetos de geração e transmissão de energia elétrica.

No início dos anos 2000, o Brasil começou a realizar leilões específicos para projetos de transmissão. Esses leilões visavam atrair investidores para a construção e operação de novas linhas de transmissão, essenciais para integrar diferentes regiões do país. Ao longo dos anos, o processo de leilões foi aprimorado para torná-lo mais eficiente e transparente.

Uma mudança significativa ocorreu no final dos anos 2000 e início dos anos 2010,

quando o Brasil começou a realizar leilões de energia de longo prazo. Nesses leilões, os investidores recebiam contratos de fornecimento de energia por um período de 20 a 30 anos. Esses contratos de longo prazo garantiam a viabilidade financeira dos projetos de transmissão, incentivando os investimentos no setor.

Além disso, o Brasil também começou a incluir projetos de energias renováveis, como energia eólica e solar, nos leilões de transmissão a partir de meados da década de 2010. Isso refletiu o compromisso do país com fontes de energia limpa e sustentável.

Durante a última década, o país continuou a realizar leilões de transmissão, incorporando atualizações regulatórias e tecnológicas para melhorar a eficiência e a confiabilidade da rede elétrica. A introdução de tecnologias avançadas, como sistemas de monitoramento em tempo real, ajudou a modernizar a infraestrutura de transmissão do Brasil.

## **1.2 Motivação**

Estudar leilões de transmissão de energia é fundamental por diversas razões. Primeiramente, esse conhecimento é essencial para o planejamento energético de um país. Compreender como os leilões são conduzidos permite aos governos e reguladores planejar a expansão da infraestrutura de transmissão de forma estratégica, garantindo que a eletricidade seja distribuída de maneira eficiente para atender à crescente demanda.

Além disso, o estudo dos leilões de transmissão é crucial para investidores e financiadores. Empresas que desejam investir em projetos de transmissão precisam entender as regras, os critérios de seleção e os contratos envolvidos nos leilões para tomar decisões sobre seus investimentos. Instituições financeiras também precisam estar familiarizadas com esses processos para avaliar os riscos e fornecer financiamento adequado para os projetos.

A eficiência energética é outra razão para estudar leilões de transmissão. Um sistema de transmissão eficiente reduz perdas de energia durante o transporte, o que é essencial para garantir que a eletricidade gerada seja entregue aos consumidores de maneira econômica e sustentável.

Além disso, os leilões de transmissão muitas vezes incluem projetos relacionados a fontes de energia renovável, como energia eólica e solar. Estudar esses leilões é importante para promover a transição para fontes de energia mais limpas e sustentáveis, alinhadas com os objetivos de redução das emissões de gases de efeito estufa e mitigação das mudanças climáticas.

Compreender como prever o resultado de um leilão de transmissão de energia é de

extrema importância para diversas partes interessadas no setor energético. Ao analisar padrões e tendências passadas, os participantes do mercado, incluindo empresas de energia e investidores, podem tomar decisões e ajustar suas estratégias de lance para aumentar suas chances de sucesso em leilões futuros. Isso não apenas reduz os riscos associados aos investimentos, mas também otimiza a alocação de recursos, melhorando o retorno sobre o investimento.

Para os órgãos reguladores e planejadores do setor energético, prever os resultados de leilões é essencial para o planejamento a longo prazo. Essas previsões permitem antecipar a capacidade futura de transmissão, facilitando o desenvolvimento sustentável da infraestrutura elétrica e garantindo um fornecimento estável de energia. Além disso, ao estimular a competição, as previsões precisas podem aumentar a eficiência do processo de leilão, resultando em benefícios tanto para os investidores quanto para os consumidores finais, por meio da redução de custos e promoção de tarifas mais competitivas. Portanto, estudar métodos de previsão nesse contexto é essencial para a tomada de decisões e o desenvolvimento saudável do mercado de energia.

### **1.3 Objetivo geral**

Esta dissertação tem como objetivo geral propor um modelo que justifique a diferença entre a Receita Anual Permitida (RAP) prevista em edital e a RAP vencedora em um leilão de transmissão de energia elétrica por meio de uma regressão linear bayesiana.

### **1.4 Objetivos específicos**

Dados o objetivo geral, esta dissertação tem como objetivos específicos:

1. Revisar a história dos leilões, com ênfase nos leilões de transmissão de energia elétrica no Brasil;
2. Apresentar a teoria dos leilões e sua relação com a teoria dos jogos;
3. Analisar os dados dos resultados históricos dos leilões de transmissão de energia (2003 a 2021);
4. Analisar as principais variáveis que podem influenciar na RAP proposta em um leilão; e
5. Propor um modelo de previsão de resultado com base no modelo de regressão linear bayesiana, sendo estudado, inicialmente, um modelo de regressão linear múltipla.



## **1.5 Conteúdo da dissertação**

Esta dissertação está dividida em 5 capítulos principais, sendo o primeiro a introdução, constituída por uma breve contextualização, pela motivação e pelos objetivos geral e específicos e pela definição da estrutura da dissertação. No segundo capítulo é apresentada a fundamentação teórica, que será explanado sobre a teoria e técnicas dos leilões, teoria dos jogos e leilões de transmissão de energia. No capítulo três é realizada uma breve revisão de outras bibliografias, a análise dos resultados dos leilões de transmissão e são apresentadas as metodologias abordadas na dissertação. No capítulo quatro são discutidos os resultados obtidos para as duas metodologias apresentadas. Por fim, no quinto capítulo serão realizadas as conclusões e considerações finais.

## 2 TEORIA DOS LEILÕES

Para se compreender a Teoria dos Leilões é de importante que antes tenha-se conhecimento sobre de onde ela surgiu: a Teoria dos Jogos. Neste capítulo será apresentada uma introdução sobre teoria dos jogos e suas classificações, teoria dos leilões e suas técnicas e leilão de transmissão de energia no Brasil e no mundo.

### 2.1 Introdução à Teoria dos Jogos

Nesta seção será apresentado um breve histórico sobre teoria dos jogos, bem como seus elementos e as regras de um jogo.

#### 2.1.1 Histórico

A Teoria dos jogos é um ramo da matemática que visa a escolha de decisões ótimas em condições de conflito, ou seja, ela busca entender as interações estratégicas entre agentes racionais em situações de tomada de decisão (SARTINI *et al.*, 2004).

Historicamente, vários autores foram precursores do que hoje se conhece como teoria dos jogos. Um dos primeiros trabalhos relacionados ao tema foi do matemático francês Antoine Augustin Cournot (1801-1877), que publicou em 1838 seu livro *Recherches sur les Principes Mathématiques de la Théorie des Richesses*, onde apresentou o modelo de duopólio que hoje leva seu nome. Alguns economistas afirmam que seu trabalho foi precursor da análise de jogos não-cooperativos (onde ocorrem situações de interação estratégica em que os participantes agem de forma independente e sem coordenação direta com os outros jogadores). (FIANI, 2009).

Outro pioneiro na teoria dos jogos foi o matemático alemão Ernst Friedrich Ferdinand Zermelo (1871-1953), que demonstrou que o conhecido jogo de xadrez sempre tinha uma solução, onde, independente da ação do outro jogador, sempre haveria uma estratégia vitoriosa. Anos mais tarde essa solução viria a ser o princípio da indução reversa, que refere-se a um método de análise de jogos sequenciais, nos quais os jogadores tomam decisões em momentos diferentes, com base nas ações passadas e nas expectativas sobre as ações futuras dos outros jogadores.

Outro grande nome lembrado na história de Teoria do Jogos é o do matemático Félix Edouard Justin Emile Borel (1871-1956), que foi o primeiro a formular o conceito de estratégia como um método de jogo. Borel acreditava que a guerra e a economia podiam ser estudadas de uma maneira semelhante (SARTINI *et al.*, 2004) (FIANI, 2009).

Poucos anos mais tarde, em 1928, o matemático húngaro-americano John Von Neumann publicou um artigo que até hoje considera-se como um marco no nascimento da Teoria dos Jogos, intitulado de “*Zur Theorie der Gesellschaftsspile*” o qual foi demonstrado que a solução para jogos de soma zero (jogos em que o ganho de um jogador representa necessariamente uma perda para o outro) ou estritamente não-cooperativos pode ser determinada através de ferramentas matemáticas (FILHO, 2017). Neste período, a teoria dos jogos expandiu-se para englobar uma ampla gama de disciplinas, incluindo economia, ciência política, biologia, psicologia e até mesmo inteligência artificial. No cerne da teoria dos jogos está a análise das escolhas feitas por indivíduos ou entidades quando seus resultados dependem não apenas de suas ações, mas também das ações dos outros.

Em 1944, com a publicação do livro intitulado “*Theory of Games and Economic Behaviour*”, John Von Neumann junto à Oskar Morgenstern, desenvolveu a análise dos jogos de soma zero, além de definir a representação de jogos de forma extensiva e discutir cooperação e formação de coalizações entre jogadores. Embora tenha sido um marco na Teoria dos Jogos, a obra tinha como limitação se restringir à jogos de soma zero, que se mostrava inadequadamente restritivo em campos de análise das interações entre indivíduos e organizações na sociedade, em particular na economia. A partir de 1950 John F. Nash Jr., John C. Harsanyi e Reinhard Selten vieram a contribuir com ferramentas mais elaboradas que permitissem analisar uma variedade maior de modelos de interação estratégica.

O economista húngaro John C. Harsanyi (1920-2000) teve sua principal contribuição para a teoria dos jogos no artigo *Games with Incomplete Information Played by "Bayesian" Players*, que foi publicado em 3 partes. Nele, Harsanyi trata do fato de que, muitas vezes, alguns jogadores dispõem de informação privilegiada em relação aos demais sobre algum elemento importante do jogo. Dessa forma, tem-se uma situação de informação assimétrica. O autor desenvolveu um modelo para tratar essa situação, ao qual denominou modelo de informação incompleta. Dessa forma, ele mostrou que o conceito de equilíbrio de Nash poderia ser estendido para os modelos de informação incompleta (FIANI, 2009).

Em 1965 o matemático e economista alemão Reinhard Selten (1930-2016), com o artigo *Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit*, realizou refinamento da noção de equilíbrio que ficou conhecido como "equilíbrio perfeito em sub-jogos", onde determinada estratégia, para ser considerada um equilíbrio perfeito em sub-jogos, tem de ser ótima considerando-se todos os possíveis desdobramentos do processo de interação estraté-

gica (FIANI, 2009). O trabalho realizado foi de grande importância em análises estratégicas, principalmente em jogos que envolvessem compromissos e ameaças, permitindo determinar quais eram ou não plausíveis.

John Forbes Nash Jr. (1928-2015) foi um matemático norte-americano cujas contribuições na teoria dos jogos e em campos relacionados tiveram um impacto significativo no desenvolvimento da economia, da ciência política e da matemática aplicada.

No artigo intitulado *Non-Cooperative Games de 1951*, Nash definiu uma noção de equilíbrio para modelos de jogos que não se restringia apenas aos jogos de soma zero. Sua contribuição foi de suma importância para o desenvolvimento da Teoria dos Jogos.

Anos mais tarde, em 1994, sua mais famosa contribuição, que ficou conhecida como o "Equilíbrio de Nash", rendeu o Prêmio Nobel de Economia. De forma geral, o Equilíbrio de Nash é um ponto em um jogo estratégico no qual nenhum jogador tem incentivo para desviar unilateralmente de sua estratégia, dado o conhecimento das escolhas dos outros jogadores. Esse conceito revolucionou a maneira como as interações estratégicas eram compreendidas, tornando-se fundamental na análise de jogos não-cooperativos.

Além da teoria dos jogos, Nash fez contribuições em áreas como a geometria diferencial e as equações diferenciais parciais, demonstrando sua ampla capacidade matemática. Ele também lecionou em diversas instituições, incluindo o Instituto de Tecnologia de Massachusetts (MIT).

### **2.1.2 Elementos da teoria dos jogos**

Ao longo dos anos, a Teoria dos Jogos tem sido aplicada para entender uma variedade de fenômenos do mundo real, como competição de empresas, negociações diplomáticas, evolução biológica, comportamento animal e até mesmo interações sociais cotidianas. Ela fornece percepções valiosas sobre como as decisões individuais podem ser influenciadas pelas ações de outros e como os agentes podem buscar otimizar seus resultados dentro do contexto de interações estratégicas.

A Teoria dos Jogos proporciona uma estrutura conceitual para modelar situações complexas de interação, que vão desde jogos competitivos de soma zero até jogos cooperativos de soma não-zero. Ela é fundamentada na ideia de que as decisões tomadas por cada agente afetam não apenas seus próprios resultados, mas também os resultados de outros agentes envolvidos. Dessa forma, a teoria dos jogos oferece ferramentas analíticas para prever e explicar

comportamentos estratégicos em contextos nos quais as escolhas de um agente influenciam diretamente as opções disponíveis para outros agentes.

A Teoria dos Jogos também revela que nem sempre os jogadores alcançam o equilíbrio de Nash, especialmente em jogos complexos com múltiplos jogadores e estratégias.

Segundo (FIANI, 2009), um jogo nada mais é do que uma representação formal que permite a análise das situações nas quais os agentes interagem entre si, agindo racionalmente. O entendimento dos elementos envolvidos nessa caracterização é de fundamental importância no estudo de teoria dos jogos. Alguns desses elementos são:

- Definição de jogo como modelo formal: sabendo que a teoria dos jogos envolve técnicas de descrição e análise, devem ser estabelecidas regras para apresentar e estudar um jogo;
- Interações: as ações de cada agente, ou jogador, afeta os demais. Em alguns casos, alguns autores também consideram os casos os quais as ações de um jogador não impactam os demais;
- Agentes: também chamado de jogador, é definido como qualquer indivíduo ou grupo deles com capacidade de decisão que interfira os demais. Os jogadores podem tanto ser indivíduos quanto organizações;
- Racionalidade: assumindo que os jogadores são racionais, supõe-se que eles empregam os meios mais adequados para tomada de ação;
- Comportamento estratégico: entende-se que cada jogador, ao tomar sua decisão, leva em consideração que os demais jogadores interagem entre si, dessa forma, sua decisão terá consequência sobre os demais.

Portanto, conforme a definição de (COLELL *et al.*, 1995), um jogo pode ser definido como uma representação formal na qual um número de indivíduos interage em uma situação de interdependência estratégica.

### **2.1.3 Regras de um jogo**

Entendendo que um jogo deve ter os elementos bem definidos para entendimento do seu modelo teórico, deve-se, primeiramente, descrever as regras universais para qualquer jogo. Dessa forma, (MOCHÓN; SáEZ, 2015) citam as seguintes regras:

- Em relação aos jogadores: saber quantos são e se seus interesses são conflitantes ou correspondentes. Sendo os jogadores agentes econômicos racionais que tomam decisões com um conjunto bem definido de ações e estratégias, assume-se que seus objetivos são

maximizar sua utilidade ou utilidade esperada. A menos que explicitamente declarado, supõe-se que os jogadores maximizam sua utilidade maximizando seus ganhos em um jogo.

- Tipo de informação: saber quais informações cada jogador possui – se são completas, simétricas ou perfeitas sobre as ações e recompensas uns dos outros; além disso, deve-se conhecer as sequências de movimento dos jogadores. A informação é o conhecimento que cada jogador tem sobre o jogo, podendo incluir o número de jogadores, o conjunto de ações de cada jogador, estratégias, recompensas e a sequência de movimentos. As informações dos jogadores podem ser perfeitas ou imperfeitas, completas ou incompletas, simétricas ou assimétricas.
  - Informação perfeita significa que nenhum movimento é simultâneo e cada jogador conhece a sequência de movimentos e para onde os jogadores se movem. Todos os jogos de movimentos simultâneos são jogos de informação imperfeita. Um jogo de informação incompleta ou assimétrica é também um jogo de informação imperfeita.
  - Em um jogo de informação incompleta, existem algumas incertezas sobre as ações dos jogadores, a sequência de movimentos do jogo ou os *payoffs*. Por exemplo, alguns jogadores escolhem seu conjunto de ações aleatoriamente. Um jogo de informações incompletas pode incluir probabilidades em alguns dos nós do jogo, onde este nó é representado por uma etapa do jogo onde um jogador toma uma decisão. Um jogo de informação incompleta é também um jogo de informação imperfeita.
  - Em um jogo de informação simétrica, os jogadores possuem os mesmos elementos em seus conjuntos de informações, incluindo a sequência do jogo, onde cada jogador escolhe uma ação, e os nós finais. Caso contrário, um jogo é chamado de “jogo de informação assimétrica”. Em jogos de informação assimétrica, os jogadores têm informações diferentes sobre os movimentos ou pagamentos uns dos outros.

Uma suposição implícita importante da Teoria dos Jogos é que a estrutura do jogo é de conhecimento comum, dessa forma os jogadores sabem quantos jogadores estão no jogo, sua sequência de movimentos, se é simultânea ou não, e o conjunto de ações ou estratégias disponíveis.

- Ações ou Estratégias: ações ou estratégias que os jogadores podem ter; interação e comunicação entre eles. Em um jogo, cada jogador tem um conjunto de ações que inclui seus possíveis movimentos ou estratégias. Os jogadores determinam suas estratégias com

base nas informações disponíveis no início do jogo e em cada momento de decisão.

- Pagamentos (*payoff*): são os resultados possíveis para cada jogador. Os pagamentos são o que os jogadores recebem no final do jogo. A natureza dos jogos é que os pagamentos diferem dependendo das ações dos jogadores. Normalmente, visualiza-se os possíveis pagamentos em uma matriz de retorno (matriz de *payoffs*).

Conforme supracitado, os modelos da teoria dos jogos descrevem a interação estratégica entre muitos jogadores, assumindo que os jogadores tomam decisões racionais para maximizar sua utilidade esperada. A interdependência das decisões dos jogadores é a base da teoria dos jogos, essas interações surgem de duas formas: sequenciais e simultâneas. A interação sequencial refere-se a cada jogador realizando uma ação em uma sequência de turnos. Durante a “vez” de um jogador, ele está ciente das ações realizadas nas rodadas anteriores. Além disso, cada um está ciente de que suas ações atuais afetarão as ações posteriores dos outros jogadores, bem como suas ações futuras durante o jogo. Já as interações simultâneas ocorrem quando os jogadores realizam ações simultaneamente, ignorando as ações atuais dos outros. É importante observar que, embora os jogadores não conheçam as ações específicas dos outros jogadores, eles estão cientes uns dos outros em jogos simultâneos.

Para analisar jogos de movimentos sequenciais e de movimentos simultâneos, a literatura apresenta dois tipos de ilustração: a matriz de pagamento e o diagrama de árvore. Uma matriz de pagamentos é uma tabela que mostra as recompensas (ou pagamentos) que os jogadores recebem com base nas combinações de escolhas que fazem em um jogo. Cada célula da matriz representa um possível resultado, indicando os ganhos (ou perdas) de cada jogador para essa combinação específica de ações. As matrizes de pagamentos são frequentemente usadas em jogos de dois jogadores, também conhecidos como jogos bimatrixiais. Uma forma de exemplificar a matriz de *payoffs* é com o Dilema do Prisioneiro, um exemplo clássico de situação em que se pode aplicar o conceito de Equilíbrio de Nash na teoria dos jogos. A matriz do dilema é representada da seguinte forma:

Tabela 3 – Matriz de *payoffs*

		Jogador 2	
		Cooperar	Trair
Jogador 1	Cooperar	R,R	S,T
	Trair	T,S	P,P

Fonte: Elaborado pelo autor.

em que:

- R (*Reward*): Recompensa por ambos os jogadores quando ambos cooperam.
- T (*Temptation*): Recompensa para o jogador que trai enquanto o outro coopera.
- S (*Sucker's Payoff*): Recompensa para o jogador que coopera enquanto o outro trai.
- P (*Punishment*): Punição para ambos os jogadores quando ambos traem.

As recompensas são organizadas na matriz de pagamentos para refletir a ordem de preferências dos jogadores. Em geral,  $R > P > S > T$ , o que significa que ambos os jogadores preferem a recompensa mais alta (R) e evitam a punição mais severa (T).

O dilema surge porque, embora ambos os jogadores tenham um incentivo individual para trair (pois isso maximizaria seus próprios resultados independentemente da ação do outro jogador), a melhor solução coletiva seria que ambos cooperassem. No entanto, a falta de comunicação e a busca individual pelo melhor resultado muitas vezes resultam em ambos os jogadores traírem, levando a um resultado sub-ótimo para ambos.

O Dilema do Prisioneiro é usado para ilustrar a tensão entre o interesse próprio e o interesse coletivo em situações de tomada de decisão, destacando a importância da coordenação e da confiança mútua para alcançar resultados melhores.

Tabela 4 – Exemplo de matriz de *payoffs*

		Jogador B	
		Cooperar	Trair
Jogador A	Cooperar	-1,-1	-3,0
	Trair	0,-3	-2,2

Fonte: Elaborado pelo autor.

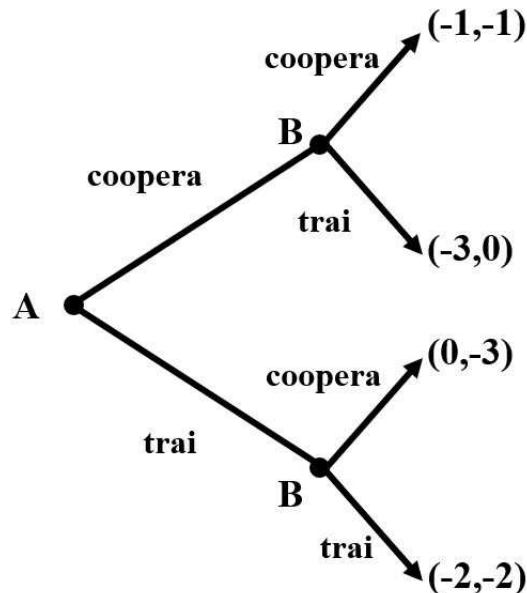
Na matriz mostrada na Tabela 4, o primeiro número em cada célula representa o pagamento de A, e o segundo número representa o pagamento de B. Por exemplo, se ambos cooperarem, cada um recebe -1. Se A cooperar e B trair, A recebe -3 e B recebe 0, onde o sinal negativo representa o fato de que o tempo de punição é algo que os participantes querem minimizar.

O diagrama de árvore é outra representação muito importante para ilustrar e analisar jogos, especialmente jogos em que os jogadores agem sequencialmente. Uma representação comum de possíveis resultados ao longo do tempo, e que se utiliza repetidamente, é chamada de “árvore”, “árvore de eventos” ou “árvore de decisão”. Tal dispositivo combina o aspecto do tempo e as conexões lógicas e sequenciais entre os tomadores de decisão e os eventos (MOCHÓN; SáEZ, 2015).



Usando o mesmo exemplo do Dilema do Prisioneiro, um diagrama de árvore seria desenhado conforme mostrado na Figura 7.

Figura 7 – Diagrama de árvore - Dilema dos Prisioneiros



Fonte: Elaborado pelo autor.

Nesse diagrama, a primeira decisão é de A (coopera ou trai), e a segunda decisão é de B (coopera ou trai). Cada ramo representa um caminho de decisões e leva aos pagamentos correspondentes para ambos os jogadores.

Tanto a matriz de pagamentos quanto o diagrama de árvore são ferramentas úteis para visualizar e analisar situações de jogo, permitindo uma compreensão mais clara das escolhas e consequências de cada jogador em uma interação estratégica.

## 2.2 Teoria dos Leilões como uma parte da Teoria dos Jogos

De acordo com (MOCHÓN; SÁEZ, 2015) um leilão é um mecanismo de mercado, operando sob regras específicas, que determina a quem um ou mais itens serão premiados e a que preço. Os leilões são de grande importância principalmente em ambientes os quais é difícil se definir um preço de mercado, pois são capazes de mensurar quanto vale uma obra de arte, um item raro de colecionador, entre outros. Dessa forma, em um ambiente de incerteza, em que a escolha do preço é uma decisão complicada, os leilões são a melhor forma de comprar e vender. Em (MENEZES; MONTEIRO, 2008), o autor explica que os leilões são frequentemente utilizados na venda de bens para os quais não existe um mercado estabelecido.

Os leilões têm uma história antiga e remontam a civilizações antigas, onde eram utilizados como um método para vender bens e serviços de forma pública e transparente. O conceito de leilões evoluiu ao longo do tempo e em diferentes culturas, mas sua essência básica de permitir que compradores compitam para adquirir itens ainda prevalece. As primeiras referências históricas registram que, lamentavelmente, os babilônios leiloavam mulheres em idade de se casar. Certas sociedades também usavam leilões para vender escravos. Os romanos faziam leilões para distribuir os espólios de guerra, que eram leiloados sob uma lança cravada no chão. Durante a Idade Média, os leilões continuaram a ser usados, muitas vezes em feiras e mercados, para vender itens como animais, produtos agrícolas e outros bens.

No século XVIII, com o crescimento do comércio e do mercado, os leilões ganharam popularidade em diversos setores, incluindo obras de arte, antiguidades, propriedades e comércio de escravos (infelizmente). Leilões de arte e antiguidades tornaram-se especialmente notáveis, com casas de leilões estabelecendo-se como intermediários para transações valiosas. Durante o século XIX, os leilões expandiram-se para incluir uma variedade de bens, como mercadorias em navios, produtos manufaturados e bens imóveis. Leilões de propriedades e terras se tornaram uma parte importante do mercado imobiliário.

Nos últimos anos, os leilões têm apresentado forte crescimento, impulsionados em grande parte pelo avanço da Internet, que surgiu como o canal ideal para conectar oferta e demanda de qualquer item, eliminando barreiras geográficas e temporais e reduzindo custos de transação. O melhor exemplo da forte penetração da web no campo dos leilões tem sido a proliferação de empresas dedicadas exclusivamente à realização de leilões online, como é o caso do *eBay*.

Hoje, os leilões são uma parte integral de muitos mercados, incluindo arte, antiguidades, imóveis, veículos, bens de consumo e muito mais. Eles podem assumir diversas formas, como leilões ao vivo, leilões online e leilões silenciosos. Além disso, os governos, cientes das vantagens dos leilões, usam esses mecanismos para alocar e definir preços de venda em ambientes de incerteza. Alguns dos bens públicos que já foram alocados por meio de leilões incluem licenças de espectro de rádio, direitos de emissão de  $CO_2$ , títulos, entre outros. Os governos são atualmente os operadores econômicos que mais recorrem aos leilões, não só para vender bens públicos, mas também para comprar bens e serviços a fornecedores que ofereçam o preço mais baixo.

## 2.3 Técnicas de Leilões

Existem vários tipos de leilões, cada um com regras específicas para determinar o vencedor e o preço pago pelo item leiloado. Segundo (MOCHÓN; SáEZ, 2015), os tipos mais comuns de leilões incluem:

- Leilão Inglês (*ascending-bid auction*): é um tipo de leilão que ocorre de forma oral, normalmente conhecido com viva-voz, onde os lances são ascendentes, ou seja, são gradualmente aumentados até que não haja mais lances e de primeiro preço, quando o valor a ser pago é o da oferta vencedor. De acordo com (KRISHNA, 2002), o leilão inglês é, talvez, o formato de leilão mais utilizado, pois possui muitos recursos atraentes; do ponto de vista dos licitantes é estrategicamente simples, cada participante só precisa de decidir se desiste ou não naquele instante;
- Leilão Holandês (*descending-bid auction*): assim como o leilão inglês, também ocorre de forma oral e é de primeiro preço, entretanto os lances são descendentes, ou seja, o preço começa alto e é gradualmente reduzido até que um licitante aceite o preço e ganhe o item.
- Leilão de Primeiro Preço Selado (*first-price sealed-bid auction*): Os licitantes apresentam seus lances em envelopes fechados. O licitante com o maior lance ganha, pagando o valor do seu lance.
- Leilão de Segundo Preço Selado (*second-price sealed-bid auction*, ou leilão de Vickrey): Similar ao primeiro preço selado, mas o vencedor paga o valor do segundo maior lance.
- Leilão de Combinatória (*combinatorial auction*): Permite aos licitantes fazer lances em múltiplos itens, possibilitando que diferentes combinações de itens sejam vendidas.
- Leilão de Valor Comum (*common-value auction*): Os licitantes têm informações assimétricas sobre o valor do item, que é o mesmo para todos, mas desconhecido. Isso cria desafios em estimar o valor real.

Os leilões de primeiro preço são mais comuns, mas também há um outro tipo de leilão onde o vencedor paga o valor do segundo maior lance. Esta modalidade de leilão é chamada leilão de segundo preço e também é conhecido como Leilão de Vickrey, uma homenagem ao economista William Vickrey (1914-1996), que realizou trabalhos essenciais na análise de leilões e recebeu Prêmio Nobel de Economia de 1996.

A teoria dos leilões investiga questões como estratégias de licitação, o efeito das informações assimétricas, o impacto das diferentes regras de leilão nos resultados e a busca pelo equilíbrio de Nash no contexto das licitações.

## 2.4 Leilões de Transmissão no Brasil

No Brasil, os leilões de transmissão de energia elétrica são realizados com o objetivo de contratar empresas ou consórcios para desenvolver, construir e operar novas instalações de transmissão de energia, bem como expandir e melhorar a infraestrutura existente. Esses leilões fazem parte da estratégia do governo para garantir a expansão da capacidade de transmissão, garantindo o suprimento de energia elétrica de forma segura e eficiente.

A Agência Nacional de Energia Elétrica (ANEEL) é o órgão responsável pela regulamentação, organização e realização dos leilões de transmissão no Brasil. O processo de realização desses leilões envolve várias etapas, iniciando pelo planejamento, onde o Ministério de Minas e Energia (MME) em conjunto com a Empresa de Pesquisa Energética (EPE) elabora um Plano de Expansão da Transmissão de Energia Elétrica (PET) que identifica as necessidades de investimento na rede de transmissão. Em seguida, é feita a definição dos Projetos, onde, com base no PET, a ANEEL seleciona e define os projetos que serão licitados nos leilões. Esses projetos são agrupados em lotes, que podem abranger diversas regiões do país. A próxima etapa consiste na publicação do edital, que é realizada pela ANEEL e contém informações detalhadas sobre os projetos, critérios de habilitação, cronograma e regras do leilão.

Nos leilões de transmissão de energia elétrica no Brasil ocorrem dois tipos de leilão: o de primeiro preço selado e o holandês. O leilão é realizado de forma pública e presencial ou por meio eletrônico. As empresas interessadas em participar apresentam lances em envelope selado para os lotes de projetos, caso haja dois lances com diferença menor que a estabelecida pelo leiloeiro, o leilão passa ser do tipo viva-voz até que haja um vencedor. O vencedor de cada lote é a empresa ou consórcio que oferece o menor valor de Receita Anual Permitida (RAP), que é a remuneração anual autorizada para a empresa pela construção e operação das instalações de transmissão. Após o leilão, a empresa vencedora assina um contrato de concessão com a ANEEL. O contrato estabelece as condições para a construção, operação e manutenção das instalações de transmissão. A empresa vencedora é responsável por desenvolver os projetos, obter as licenças ambientais necessárias, construir as instalações de transmissão e operá-las conforme os prazos e condições estabelecidos no contrato.

Os leilões de transmissão no Brasil visam atrair investidores e promover a competição entre as empresas interessadas em desenvolver a infraestrutura de transmissão. Essa abordagem ajuda a garantir a eficiência na expansão da rede elétrica e a manter a segurança e confiabilidade do sistema elétrico do país.

No Brasil, o desenvolvimento do sistema de transmissão de energia elétrica está intimamente ligado à história da eletrificação do país. A construção de redes de transmissão elétrica foi um marco importante para o desenvolvimento econômico e social do Brasil ao longo do século XX. No início do século XX, diversas cidades brasileiras começaram a estabelecer sistemas elétricos locais para iluminação pública e uso industrial. A criação da *Light and Power*, uma empresa canadense, marcou o início da eletrificação no Rio de Janeiro em 1904.

Entre as décadas de 1940 e 1950 ocorreu um aumento significativo na construção de usinas hidrelétricas e na expansão das redes de transmissão para atender à crescente demanda. A criação do Conselho Nacional de Águas e Energia Elétrica (Conselho Nacional de Águas e Energia Elétrica (CNAEE)) em 1948 consolidou a regulação do setor elétrico.

A década de 1960 foi marcada pelo início da construção de grandes usinas, como Itaipu e a expansão do sistema de transmissão para integrar essas novas fontes de energia. A Eletrobras, empresa estatal (até 2022) criada em 1962, desempenhou um papel fundamental no planejamento e desenvolvimento do setor elétrico no Brasil. A criação da ANEEL em 1997 marcou uma mudança significativa na regulação do setor elétrico brasileiro. A ANEEL é responsável pela regulação, fiscalização e controle dos serviços e instalações de energia elétrica no Brasil e também pela realização de leilões para concessões de transmissão.

A entrada do setor privado no final dos anos 1990 e início dos anos 2000 levou a um aumento significativo nos investimentos em novas usinas e infraestrutura de transmissão. A expansão da matriz energética incluiu um foco crescente em energias renováveis, como a eólica e a solar, levando a investimentos em novas linhas de transmissão para integrar essas fontes ao sistema.

Para que o sistema de transmissão funcione de forma segura e confiável, é preciso que haja gestão e planejamento, onde o Operador Nacional do Sistema Elétrico (ONS) é responsável pelo gerenciamento do sistema elétrico brasileiro, monitorando o consumo, a geração e a transmissão em tempo real. Além disso, a Empresa de Pesquisa Energética (EPE) é encarregada do planejamento energético de longo prazo.

Para conceder contratos de construção e operação de novas linhas de transmissão para empresas privadas são realizados os leilões, que incentivam investimentos em infraestrutura de transmissão. Além disso, as linhas de transmissão e as subestações são regularmente mantidas e modernizadas para garantir a segurança e a eficiência do sistema. Isso inclui a substituição de equipamentos antigos, a implementação de tecnologias de automação e a expansão da capacidade

para atender à crescente demanda

Vale ressaltar que o sistema de transmissão de energia elétrica no Brasil é regulamentado pela ANEEL, que estabelece normas, regula as tarifas e garante a qualidade e a segurança do serviço elétrico para os consumidores brasileiros.

## 2.5 Edital do Leilão

A concessão de transmissão de energia elétrica no Brasil segue um processo regulamentado pela ANEEL. Inicialmente é realizado um processo de identificação de necessidades, onde o Ministério de Minas e Energia e a Empresa de Pesquisa Energética identificam as necessidades de expansão da rede de transmissão com base nos estudos de planejamento energético de longo prazo. Em seguida é realizada a elaboração do edital, no qual, com base nas necessidades identificadas, a ANEEL elabora uma documentação, que é disponibilizada no site da agência (ANEEL, 2023b) (contém os editais e documentações desde o ano de 1999), que inclui informações detalhadas sobre os lotes de transmissão, critérios de habilitação, regras do leilão e compromissos de investimentos, como mostrado no exemplo do edital do leilão nº 1/2023 na Figura 8.

Figura 8 – Edital Nº 1/2023-ANEEL



Fonte: (ANEEL, 2023c)

Os editais de concessão de transmissão de energia elétrica são documentos detalhados que estabelecem as regras, critérios e condições para a participação de empresas interessadas em concorrer aos leilões de concessão. Eles são cuidadosamente elaborados para garantir um processo competitivo, transparente e eficiente, fornecendo informações detalhadas para orientar as empresas interessadas durante todo o processo de participação no leilão e subsequente

execução da concessão. Na Figura 9 é possível verificar o sumário do edital N° 1/2023.

Figura 9 – Sumário do Edital N° 1/2023-ANEEL

Número: 48577.000530/2023-00



**Edital do Leilão nº 1/2023-ANEEL (Leilão de Transmissão) - Processo nº 48500.007497/2022-04**

---

**SUMÁRIO:**

1. INFORMAÇÕES PRELIMINARES.....	4
2. PARTICIPAÇÃO.....	6
3. OBJETO.....	8
4. DIREITOS E OBRIGAÇÕES GERAIS.....	10
5. RECEITA ANUAL PERMITIDA PELA PRESTAÇÃO DO SERVIÇO DE TRANSMISSÃO .....	15
6. FORMA DE APRESENTAÇÃO DOS DOCUMENTOS .....	16
7. INSCRIÇÃO PARA PARTICIPAR NO LEILÃO .....	17
8. GARANTIA DE PROPOSTA .....	18
9. SESSÃO PÚBLICA DO LEILÃO.....	21
10. HABILITAÇÃO.....	24
11. GARANTIA DE FIEL CUMPRIMENTO DO CONTRATO .....	32
12. ADJUDICAÇÃO DO OBJETO.....	35
13. PENALIDADES .....	35
14. ASSINATURA DO CONTRATO DE CONCESSÃO .....	40
15. RECURSOS .....	42
16. DISPOSIÇÕES FINAIS .....	43
17. CRONOGRAMA DO LEILÃO.....	44

Fonte: (ANEEL, 2023c)

Os editais dos leilões contêm informações cruciais para os investidores, dentre elas:

- Descrição dos lotes: os editais apresentam uma descrição detalhada dos lotes de transmissão disponíveis para concessão, incluindo informações sobre a localização, capacidade de transmissão, tecnologia a ser utilizada, entre outros;
- Critérios de participação: estabelecem os requisitos que as empresas devem atender para participar do leilão, incluindo aspectos técnicos, financeiros, e experiência prévia no setor de transmissão;
- Garantias: especificam as garantias que as empresas interessadas devem apresentar para participar do leilão, como garantias de oferta e de performance;
- Regras do leilão: detalham as regras do processo de leilão, incluindo a forma de apresentação das propostas, critérios de seleção, lances mínimos, entre outros;
- Compromissos de investimento: estabelecem os compromissos que a empresa vencedora

deve cumprir, como prazos para construção das instalações de transmissão e investimentos mínimos necessários;

- Tarifas e receitas: especificam as tarifas máximas que podem ser cobradas dos usuários para remunerar os investimentos, bem como as regras para revisões tarifárias ao longo do período de concessão;
- Aspectos ambientais e licenciamento: incluem as obrigações e requisitos relacionados ao licenciamento ambiental das instalações de transmissão, garantindo a conformidade com as leis e regulamentações ambientais;
- Prazos e cronograma: estabelecem os prazos para entrega de documentos, participação no leilão, assinatura de contratos e início das operações comerciais das instalações de transmissão;
- Penalidades: especificam as penalidades em caso de descumprimento dos compromissos estabelecidos no edital ou no contrato de concessão.



### 3 ANÁLISE DOS LEILÕES DE TRANSMISSÃO

Nesta seção serão apresentadas uma revisão bibliográfica com referências de trabalhos realizados no tema bem como a metodologia estudada e resultados obtidos, as variáveis do modelo utilizado na dissertação e, por fim, os equacionamentos para análise do modelo escolhido.

#### 3.1 Revisão Bibliográfica

A análise das principais variáveis possíveis de ter relação com o valor final da proposta ganhadora em um leilão de transmissão foi tema de discussão para diversos outros autores. Em (CEZARIO *et al.*, 2009) foram analisados os leilões entre os anos 2000 a 2007 e o modelo estudado foi utilizado um modelo de regressão por mínimos quadrados. Os autores utilizaram como variável dependente a RAP do proponente vencedor e como variáveis independentes utilizaram a RAP proposta em edital, o número de competidores, a extensão da linha de transmissão e uma variável *dummy* para lances inferiores à 48 milhões (variável de escolha dos autores). Além disso, os autores testaram com variações de variável dependente, incluindo o investimento e o prazo de execução. Após realizados os testes no modelo, os autores concluíram que as variáveis independentes escolhidos explicavam quase 60% do modelo e que a variável mais significativa foi a número de competidores.

Em (RASO, 2013) o autor analisou dados no período de 2003 a 2013, com informações de cerca de 35 leilões, constituídos de 222 lotes no total. Foram analisados a expansão das linhas de transmissão ao longo dos anos, a quantidade de leilões por lote, o percentual da participação de empresas por consórcio, podendo ser públicas ou privadas, e de empresas com participação isolada. Além disso, a participação estrangeira e a média de deságio por ano e a relação entre o número de participantes e o deságio. O autor realizou uma análise econométrica com as variáveis deságio, lances, participantes, proponentes e extensão, validando a também a correlação entre as variáveis. Por fim, por meio de um modelo de regressão linear, o autor concluiu que a única variável que explica o deságio é a variável proponentes.

No trabalho (NASCIMENTO, 2012) o autor realizou um estudo com resultados de leilões ocorridos entre os anos de 1999 e 2010, totalizando em 136 lotes leiloados com 577 propostas. A variável dependente utilizada foi o deságio (razão entre a diferença do RAP do edital e do RAP proposto pelo RAP do edital) e suas variáveis independentes são: investimento,

número de participantes, linha de transmissão (variável com valor 1 se predominar linha de transmissão no lote), subestação (variável com valor 1 se predominar subestação no lote), consórcio, participação estatal, participação estrangeira e região. Foram analisadas a média anual do deságio, o número de lotes e o deságio por região, o número de proponentes com participação isolada ou em consórcio, com e sem participação estatal e com e sem participação estrangeira. O autor utilizou modelo de regressão linear múltipla e analisou a significância das variáveis no modelo, concluindo que as mais significativas para o modelo foram as variáveis investimento, número de participantes, linha de transmissão, subestação, estatal, estrangeira e região Nordeste.

Em (CAZZARO, 2017), o autor busca analisar a participação dos agentes investidores no mercado de transmissão de energia elétrica, com foco no perfil dos que venceram o leilão e nas seguintes variáveis: capital público ou privado, nacional ou estrangeiro. Neste trabalho foram analisados dados de 1999 a 2017 e nele o autor fez análises percentuais da composição de cada leilão quanto à composição público ou privada, participação estrangeira e consórcios.

O trabalho de (CASTRO; MARTINI, 2021) tem como objetivo identificar as causas que explicam um deságio médio do leilão ocorrido em dezembro de 2020. O autor afirma que os deságios refletem a qualidade dos contratos e que podem ser destacados três variáveis que permitem explicar os deságios no leilão: a primeira é a externalidade para alguns concorrentes, nos casos de empresas que detêm conhecimento da região do empreendimento; a segunda é a estratégia de ampliação do portfólio; e a terceira variável refere-se às empresas com pouca atuação no mercado de transmissão.

No trabalho realizado em (OLIVEIRA *et al.*, 2023) os autores analisaram os resultados dos leilões ocorridos entre os anos de 2003 e 2021. Utilizou-se um modelo de regressão linear múltipla, onde a variável dependente era a diferença entre a RAP do edital pela RAP proposta e foram levadas em consideração as variáveis independentes o número de competidores, o número de empreendimentos, a extensão das linhas de transmissão e o prazo para construção. Da mesma forma que os demais trabalhos, verificou-se a significância das variáveis e como elas explicam o resultado da variável dependente. Os resultados deste trabalho mostraram que as variáveis independentes explicavam aproximadamente 45,8% da variabilidade do valor do lance no leilão e que as variáveis mais relevantes na determinação do deságio foram número de competidores, extensão das linhas de transmissão e prazo para execução do empreendimento.

Em (FRACASSO, 2019) o autor realiza um estudo nas principais causas dos valores de deságio nos leilões ocorridos entre os anos de 2011 e 2018. A relação entre as variáveis foi

estabelecida através de um modelo de regressão linear por meio do método dos mínimos quadrados ordinários. A variável dependente do modelo foi o deságio e as variáveis independentes era a o investimento, rentabilidade, taxa SELIC, taxa Risco Brasil, prazo, número de propostas nos lotes e licenciamento Ibama. Neste trabalho foram analisados os investimentos realizados pelo setor público e pelo setor privado e a quantidade de lotes, investimento total, média de ofertantes, taxa Risco Brasil, taxa SELIC e rentabilidade por lote. No estudo realizado, o modelo utilizado mostrou que as variáveis que mais afetam o deságios são investimento, taxa SELIC e rentabilidade.

Após estudos nos trabalhos supracitados e em outros que buscavam encontrar os principais fatores que influenciavam nos resultados dos leilões, observou-se que grande parte levava em consideração praticamente as mesmas variáveis, havendo pouca mudança também quanto à metodologia utilizada. Dessa forma, para esta dissertação foram levantadas algumas das variáveis discutidas em outros trabalhos e também, visando obter melhores resultados, outras variáveis que pudessem ter relação com o valor final da diferença entre a RAP do edital e a RAP proposta.

### **3.2 Variáveis do Modelo**

Para que se obter as variáveis que pudessem ter relação com o valor final da diferença entre a RAP do edital e RAP proposta, foi construído um diagrama para entender quais as características do lote a ser leiloado a quais as características do proponente, conforme mostrado na Figura 10.

Após realizados estudos dos editais dos leilões de transmissão de energia elétrica e sabendo que eles são coordenados pela ANEEL, acessou-se a base de dados históricos dos leilões disponível site da agência ANEEL (2023a), onde foi possível obter uma planilha no formato *xlsx* que continha as seguintes informações no período de 1999 e 2021:

- número do leilão: no formato número/ano;
- número do lote: número do lote do leilão em questão, podendo ser classificado numericamente ou por letras;
- empreendimentos: qual tipo de empreendimento do lote – podendo ser linha de transmissão ou trechos dessas ou subestação;
- UF: estado(s) no qual ocorrerá a instalação da linha de transmissão e/ou da subestação;
- extensão (km): extensão em quilômetros da linha de transmissão que será construída;

- prazo (meses): período em meses para execução da obra;
- RAP - Receita Anual Permitida (R\$) – Edital e Proposta: RAP conforme previsto em edital e RAP do proponente ganhador;
- Deságio (%): razão entre a diferença do RAP do edital pelo RAP proposto pelo RAP edital;
- Concessionária: proponente que arrematou o lote, podendo ser participação do tipo consórcio ou isolada.

Os dados obtidos no site da ANEEL não estavam com qualidade apropriada para ser trabalhada, contendo algumas características que necessitam de um refino para ser utilizada, entre elas, pode-se citar: muitas células mescladas; falta de padrão na formatação de número (por exemplo, para o mesmo parâmetro, hora aparte decimal era separada por ponto, hora separada por vírgula); depois de cada leilão havia uma linha com totais na planilha; alguns lotes não continham informações de um ou mais parâmetros; hora o deságio estava em percentual, hora em decimal; entre outros.

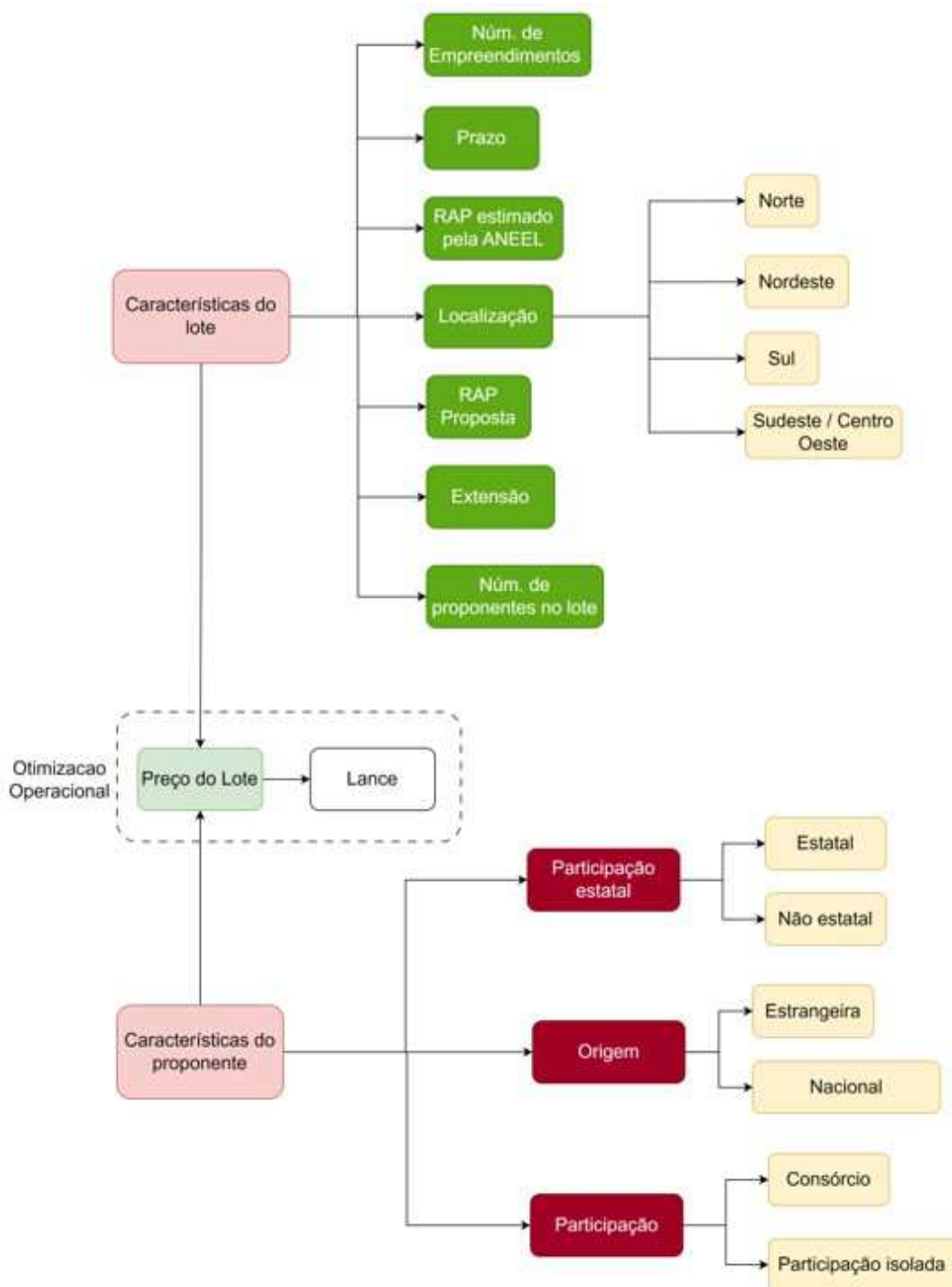
Dessa forma, após intensiva análise dos dados, foi realizado um tratamento nas informações para que fosse possível trabalhar com elas. Entre as ações realizadas, pode-se citar:

- Desassociar as células que estavam mescladas;
- Padronização da coluna "Extensão km" para número sem casas decimais;
- Mudança das unidades de "RAP - Receita Anual Permitida (R\$) - Edital" e "RAP - Receita Anual Permitida (R\$) - Proposta" de reais para milhões de reais;
- Padronização da coluna "Deságio %" com formatação em percentual;
- Exclusão de lotes dos leilões que estivessem faltando alguma das informações dos parâmetros;
- Em cada dado de empreendimento a informação foi estratificada e foram contabilizados o número de empreendimentos, sendo essas linhas de transmissão ou subestação, de cada lote;
- A partir da informação da UF do lote, a informação foi classificada quanto à região. Os estados foram classificados de acordo com os subsistemas do SIN;

A base de dados era formada pelas principais informações dos editais bem como o valor pelo qual a proposta vencedora arrematou o lote e quem foi o vencedor do leilão. Entretanto, não havia informações acerca dos demais competidores em cada leilão. Dessa forma, para obter as informações dos demais participantes e suas respectivas propostas foram acessados os dados

do site da B3 (2023). A B3, sigla para Brasil, Bolsa e Balcão, é uma das principais empresas de infraestrutura de mercado financeiro no mundo, seu formato atual é resultante da fusão entre a Bolsa de Valores, Mercadorias e Futuros de São Paulo (BM&FBOVESPA) com a Central de Custódia e de Liquidação Financeira de Títulos (CETIP).

Figura 10 – Diagrama causal - Preço do Lote



Fonte: Elaborado pelo autor.

Por meio de uma programação em @Python a partir da URL da página da B3 (2023),

foi possível extrair um banco de dados com as informações de todos os participantes e seus respectivos lances em cada lote entre os anos de 2003 e 2022. De posse dos dados obtidos, a etapa seguinte consistiu em tratar os dados para que eles pudessem ser melhor analisados. Dessa forma, foram realizadas os seguintes passos:

1. Para que as informações tanto da ANEEL quanto da B3 pudessem ser mescladas, foram selecionados apenas os dados dos mesmos períodos, dessa forma, a base de dados foi tratada com informações dos leilões ocorridas entre os anos de 2003 e 2021;
2. Outra tratativa dada foi em relação à exclusão de informações inconsistentes. Foram desconsiderados os lotes que estavam sem prazo de execução, lotes os quais o leilão não aconteceu ou lotes que estavam sem informações da ANEEL;
3. Os participantes dos leilões foram classificados de acordo com sua origem: se eram consórcio ou participação isolada, se tinham participação estrangeira ou não e se tinha participação estatal ou não. Para as variáveis não numéricas foram classificadas como variáveis *dummy*, que é uma variável que assume valores de "zero" ou "um" indicando a ausência ou presença de qualidades ou atributos.

As informações dos dois bancos de dados foram agregadas por meio da criação de uma chave que mesclava o ano do leilão ao número de lote. Com o auxílio do software PowerBI foi montada uma única base de dados no formato *.xlsx* e foram realizadas as primeiras análises gráficas. Esta base é formada por 343 observações (lotes dos leilões) no período de 2003 a 2021. Após a tratativa dos dados, pôde-se chegar às seguintes variáveis que seriam utilizadas no modelo:

- num\_comp: número de competidores em cada lote;
- emp: número de empreendimentos por lote;
- ext: extensão da linha de transmissão;
- prz: prazo de execução;
- nrt: caso a obra fosse realizada na região norte;
- ne: caso a obra fosse realizada na região nordeste;
- sul: caso a obra fosse realizada na região sul;
- se\_co: caso a obra fosse realizada na região sudeste/centro-oeste;
- estat: se havia participação estatal;
- estrang: se havia participação estrangeira;
- cons: se o proponente era consórcio

- dif: a diferença entre o RAP do edital e o RAP proposto (em milhões).

A escolha das variáveis para estudos desta dissertação foi baseada em outros trabalhos da literatura que sugeriam que algumas dessas variáveis pudessem ter correlação com o valor final da diferença entre a RAP do edital e a RAP proposta pelo proponente. Dessa forma, foram levantados o maior número possível de variáveis, em seguida, as mesmas foram analisadas se seriam relevantes para o modelo.

### **3.3 Análise de modelos estatísticos**

Após definidas as variáveis, foram utilizados dois modelos: inicialmente um Modelo de Regressão Linear Múltipla que foi posteriormente substituído por um modelo de Regressão Linear Bayesiana.

#### ***3.3.1 Modelo de Regressão Linear Múltipla***

Um modelo de regressão linear é utilizado quando há uma relação linear entre duas variáveis, ou seja, quando uma variável dependente (ou resposta) pode ser prevista com base em uma ou mais variáveis independentes (ou preditoras). A regressão linear é uma técnica estatística amplamente utilizada em várias áreas, incluindo economia, finanças, ciências sociais, ciências naturais e engenharia.

Este tipo de modelo pode ser usado para previsão, quando se deseja prever um valor numérico com base em dados históricos ou outras variáveis conhecidas, como para prever o preço de uma casa com base em sua área, número de quartos, localização, etc; pode ser utilizado também para se estabelecer relacionamentos causais, quando se quer entender o relacionamento causal entre variáveis, por exemplo, entender como a quantidade de estudo afeta as notas dos alunos; para identificar tendências e padrões ao longo do tempo, como analisar o crescimento populacional ao longo dos anos; em algumas situações, pode-se utilizar regressão linear para eliminar variáveis desnecessárias, identificando quais variáveis independentes têm um impacto significativo na variável dependente.

Na engenharia e manufatura, um modelo de regressão linear pode ser usado para controle de processos para entender a relação entre variáveis de processo e a qualidade do produto. Em negócios, pode ser usada para prever vendas com base em variáveis como publicidade, preço do produto, promoções, entre outros. Já em economia e ciências sociais, a regressão linear pode

ser usada para avaliar o impacto de políticas públicas em diferentes variáveis, como taxa de emprego, taxa de criminalidade, e em finanças, pode ser usada para entender como diferentes variáveis (como taxas de juros, inflação, etc.) afetam o valor dos investimentos. Em suma, a regressão linear é uma ferramenta versátil que pode ser aplicada em uma variedade de contextos quando há uma relação linear entre variáveis e se deseja entender, prever ou otimizar um resultado numérico com base nessas variáveis.

Em um modelo de regressão linear simples, tem-se uma variável dependente (ou resposta) que é predita por uma única variável independente. No entanto, em situações mais complexas, pode haver mais de uma variável independente que influencia a variável dependente. Quando há várias variáveis independentes, é chamado de modelo de regressão linear múltipla.

Para um modelo de regressão linear múltipla com  $n$  variáveis independentes ( $X_1, X_2, \dots, X_n$ ) e uma variável dependente ( $Y$ ), o modelo é representado pela Equação 3.1 (CHEIN, 2019).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (3.1)$$

Neste equacionamento, tem se que:

- $Y$  é a variável dependente que se quer prever;
- $X_1, X_2, \dots, X_n$  são as variáveis independentes;
- $\beta_0$  é o intercepto da linha de melhor ajuste no eixo  $y$  (quando todas as variáveis independentes são 0);
- $\beta_1, \beta_2, \dots, \beta_n$  são os coeficientes que representam a mudança esperada em  $Y$  para uma mudança de uma unidade nas variáveis independentes correspondentes;
- $\varepsilon$  é o erro, que representa a diferença entre o valor observado de  $Y$  e o valor previsto pelo modelo.

Os coeficientes  $\beta_0, \beta_1, \beta_2, \dots, \beta_n$  são estimados a partir dos dados usando técnicas como o Método dos Mínimos Quadrados, que minimiza a soma dos quadrados das diferenças entre os valores observados e os valores preditos pelo modelo.

Cada coeficiente  $\beta_i$  indica como a variável dependente  $Y$  muda quando a variável independente correspondente  $X_i$  é aumentada em uma unidade, mantendo todas as outras variáveis independentes constantes.

O modelo de regressão linear múltipla é uma ferramenta amplamente utilizada para análise de dados quando há múltiplas variáveis que podem influenciar o resultado desejado. Ele



permite entender as relações complexas entre várias variáveis e fazer previsões com base nessas relações.

Inicialmente optou-se por utilizar um modelo de regressão linear múltipla para análise da variáveis e modelagem preditiva do resultados dos leilões. Foram analisadas correlações entre as variáveis calculada por meio do coeficiente de Pearson, conforme a Equação 3.2 (CHEIN, 2019).

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \times \sigma_Y} \quad (3.2)$$

Em que:

- $\text{cov}(X, Y)$ : representa a covariância entre  $X$  e  $Y$ ;
- $\sigma_X$  e  $\sigma_Y$ : representam os desvios padrão de  $X$  e  $Y$ , respectivamente.

Foi realizado um escalonamento nos dados, de modo que a média fosse igual a zero e o desvio padrão igual a um, de modo assegurar que as variáveis tenham a mesma escala e evitar que alguma delas exerça maior influência no modelo de regressão. Após o escalonamento, foi verificada a correlação entre as variáveis independentes e a saída.

O coeficiente de correlação de Pearson varia de -1 a 1. Um valor próximo de 1 indica uma correlação positiva perfeita, -1 indica uma correlação negativa perfeita, e 0 indica ausência de correlação linear. Quanto mais próximo o valor estiver de 1 (ou -1), mais forte é a relação linear entre as variáveis dependentes.

Além disso, foram calculados covariância e desvios padrões das variáveis. A correlação entre variáveis é uma medida estatística que descreve a relação entre duas ou mais variáveis, ou seja, ela indica como as variáveis tendem a se comportar em relação umas às outras. A correlação não implica causalidade (ou seja, uma variável causando a outra), mas mostra se e como as variáveis mudam juntas.

A covariância entre  $X$  e  $Y$  é calculada conforme a Equação 3.3 (MONTGOMERY; RUNGER, 2009).

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n} \quad (3.3)$$

Em que:

- $X_i$  e  $Y_i$  são os valores individuais de  $X$  e  $Y$ , respectivamente.
- $\bar{X}$  e  $\bar{Y}$  são as médias de  $X$  e  $Y$ , respectivamente.

–  $n$  é o número total de observações.

A covariância é uma medida que indica como duas variáveis aleatórias mudam juntas. Ela mede a variabilidade conjunta de duas variáveis. Uma covariância positiva indica que as variáveis tendem a aumentar ou diminuir juntas, enquanto uma covariância negativa indica que uma variável aumenta quando a outra diminui. A covariância não é padronizada e é expressa nas unidades originais das variáveis. Portanto, os valores de covariância podem variar amplamente, tornando difícil comparar a força da relação entre diferentes pares de variáveis. Ela pode variar de  $-\infty$  a  $+\infty$ .

O desvio padrão de uma variável  $X$  é calculado de acordo com a Equação 3.4 (CHEIN, 2019).

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \quad (3.4)$$

O desvio padrão é uma medida estatística que quantifica a dispersão ou variabilidade dos valores em um conjunto de dados. Em outras palavras, ele indica o quanto os valores em um conjunto de dados se afastam, em média, da média desse conjunto. Quanto maior o desvio padrão, mais dispersos os valores estão ao redor da média. Em outras palavras, há uma maior variabilidade nos dados, e quanto menor o desvio padrão, mais próximos os valores estão da média, indicando menor variabilidade.

Para calcular os coeficientes do modelo foi utilizado o Método dos Mínimos Quadrados, de acordo com a Equação 3.5 (CHEIN, 2019).

$$J(\beta_0, \beta_1, \dots, \beta_n) = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 \times X_{i1} + \beta_2 \times X_{i2} + \dots + \beta_n \times X_{in}))^2 \quad (3.5)$$

Em que  $n$  é o número de observações nos dados.

Para encontrar os coeficientes que minimizam  $J$ , deriva-se  $J$  em relação a  $\beta_0, \beta_1, \dots, \beta_n$  e iguala-se as derivadas a zero para encontrar os valores que minimizam a função de custo. Para um modelo de regressão linear múltipla, a solução analítica para encontrar os coeficientes é dada pelas equações normais:

$$X^T X \beta = X^T Y \quad (3.6)$$

Em que  $X$  é a matriz de variáveis independentes (com uma coluna unitária para o termo de intercepto),  $Y$  é o vetor da variável dependente, e  $\beta$  é o vetor de coeficientes.

O resultado final são os coeficientes  $(\beta_0, \beta_1, \dots, \beta_n)$  que minimizam a soma dos quadrados dos resíduos, permitindo assim uma melhor representação dos dados pela equação de regressão linear múltipla.

O erro padrão de uma estimativa (como o coeficiente de regressão) é calculado de acordo com a Equação 3.7 (CHEIN, 2019).

$$SE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 2}} \quad (3.7)$$

Em que:

- $y_i$ : são os valores reais;
- $\hat{y}_i$ : são os valores preditos pelo modelo;
- $n$ : é o número de observações.

Em um modelo de regressão linear múltiplo, que envolve duas ou mais variáveis independentes e uma variável dependente, várias estatísticas são calculadas para avaliar o ajuste, a precisão e a significância do modelo. Além das estatísticas citadas, calcula-se também:

#### 1. Erro Padrão dos Coeficientes:

O Erro Padrão da Estimativa (SE ou  $\lambda$ ) em um modelo de regressão linear é uma medida da dispersão dos pontos de dados ao redor da linha de regressão. Quanto menor o erro, melhor o modelo se ajusta aos dados e mais precisas são as estimativas dos coeficientes. Conforme mostrado anteriormente, o erro padrão é calculado de acordo com a 3.7 (CHEIN, 2019).

#### 2. Teste F Global:

O teste F global é usado para verificar a significância global do modelo de regressão. Ele compara o ajuste do modelo de regressão com o ajuste de um modelo nulo (um modelo que não inclui nenhuma variável independente) para determinar se há evidências suficientes para concluir que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variável dependente. O cálculo do teste F é dado pela Equação 3.8 (CHEIN, 2019).

$$F = \frac{\frac{SSR}{df_{reg}}}{\frac{SSE}{df_{res}}} \quad (3.8)$$

Os cálculos de Soma dos Quadrados de Regressão (SSR) e Soma dos Quadrados dos Resíduos (SSE) são dados pelas Equações 3.9 e 3.10, respectivamente.

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (3.9)$$

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.10)$$

Onde  $\hat{y}_i$  são os valores previstos pelo modelo de regressão.

Se a estatística F for grande e o valor-p associado for pequeno, isso sugere que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variável dependente, indicando a significância global do modelo de regressão.

### 3. Estatística $t$ :

Os testes  $t$  são usados para avaliar a significância estatística dos coeficientes de regressão individuais. Os valores-p associados indicam se cada coeficiente é estatisticamente significativo. A estatística  $t$  é utilizada para testar a hipótese nula de que o coeficiente correspondente é igual a zero, em relação à hipótese alternativa de que é diferente de zero, levando em consideração os outros preditores do modelo (MONTGOMERY; RUNGER, 2009). Este parâmetro é a razão entre o coeficiente da variável e seu erro padrão, dessa forma, os maiores valores (positivos ou negativos), indicam que a estimativa é relativamente distante do valor nulo e indicam que essas variáveis são importantes para a previsão da variável dependente. A estatística  $t$  para um coeficiente de regressão em um modelo de regressão linear simples é calculada conforme a equação 3.11 (CHEIN, 2019).

$$t = \frac{\text{Coeficiente de Regressão}}{SE(\text{Coeficiente de Regressão})} \quad (3.11)$$

### 4. Valor $p$ :

O valor  $p$  está associado à estatística  $t$  e é usado para testar a hipótese nula de que o coeficiente de regressão é igual a zero, o que significa que não tem efeito significativo no modelo. O valor  $p$  é calculado usando a distribuição  $t$ . Um valor-p baixo ( $< 0,05$ ) indica que a hipótese nula pode ser rejeitada, o que sugere que o preditor com um valor-p baixo

provavelmente terá um impacto significativo no modelo, já que as alterações nesse preditor estão relacionadas a mudanças na variável resposta (CHEIN, 2019).

Esses cálculos são fundamentais na análise de regressão para avaliar a significância estatística dos coeficientes estimados.

Para calcular o valor  $p$ , primeiro formula-se Hipótese Nula ( $H_0$ ) e a Hipótese Alternativa ( $H_1$ ):

- $H_0$ : a variável independente não tem efeito significativo sobre a variável dependente (o coeficiente é igual a zero);
- $H_1$ : a variável independente tem um efeito significativo sobre a variável dependente (o coeficiente não é igual a zero).

Em seguida, realiza-se o teste estatístico, o qual está associado com o coeficiente da variável independente que segue uma distribuição  $t$  de Student (ou uma distribuição normal em grandes amostras) quando a hipótese nula é verdadeira. Depois é verificado o número de graus de liberdade está relacionado ao tamanho da amostra e ao número de variáveis independentes no modelo.

Por fim, encontra-se o valor  $p$  associado ao  $t$ -valor, utilizando uma tabela  $t$  de *Student* ou um software estatístico para encontrar o valor  $p$  associado ao  $t$ -valor e aos graus de liberdade. Alternativamente, a maioria dos softwares estatísticos fornece essa informação diretamente ao executar uma regressão. Um valor  $\alpha$  é escolhido antecipadamente (geralmente 0,05 ou 0,01) como o limiar de significância. Se o valor  $p$  calculado for menor que  $\alpha$ , você rejeita a hipótese nula em favor da hipótese alternativa. Se o valor  $p$  é menor que o valor  $\alpha$  escolhido, isso sugere que o coeficiente é estatisticamente significativo, o que significa que há evidências suficientes para afirmar que a variável independente tem um efeito significativo sobre a variável dependente.

##### 5. Raiz quadrada do erro quadrático médio (RMSE):

Estima o desvio padrão da distribuição do erro. A Raiz Quadrada do Erro Quadrático Médio (RMSE) é uma medida comum usada para avaliar a precisão de um modelo de previsão. É calculada como a raiz quadrada da média dos quadrados dos erros entre os valores previstos pelo modelo e os valores reais. A fórmula para calcular o RMSE é dada pela Equação 3.12 (PAULA, 2004).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.12)$$

Em que:

- $n$ : é o número total de observações nos dados.
- $y_i$ : é o valor real da variável dependente para a  $i$ -ésima observação.
- $\hat{y}_i$ : é o valor previsto pelo modelo para a  $i$ -ésima observação.

Para cada observação, calcula-se o resíduo, que é a diferença entre o valor real  $y_i$  e o valor previsto  $\hat{y}_i$  pelo modelo. Os resíduos são usados para verificar as suposições do modelo, incluindo a homocedasticidade (variância constante dos erros) e a normalidade dos erros. O cálculo de resíduos se dá conforme mostrado na Equação 3.13.

$$e_i = y_i - \hat{y}_i \quad (3.13)$$

Eleva-se a 3.13 ao quadrado para eliminar o sinal e enfatizar os erros grandes e calcula-se a média dos quadrados dos erros elevados ao quadrado:  $\frac{1}{n} \sum_{i=1}^n e_i^2$  e, finalmente, calcula-se a raiz quadrada do resultado da etapa anterior para obter o RMSE.

O RMSE fornece uma medida do erro médio entre os valores previstos e os valores reais. Quanto menor o RMSE, melhor é o desempenho do modelo em termos de previsão.

## 6. Coeficiente de Determinação $R^2$ :

O coeficiente de determinação  $R^2$  é uma medida estatística que representa a proporção da variabilidade na variável dependente que é explicada pelas variáveis independentes em um modelo de regressão.  $R^2$  varia de 0 a 1, onde 0 indica que o modelo não explica nenhuma variabilidade e 1 indica que o modelo explica toda a variabilidade. Para calcular  $R^2$ , são realizados os seguintes passos:

- Soma Total os Quadrados (SST), conforme a Equação 3.14 (CHEIN, 2019).

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (3.14)$$

onde  $y_i$  são os valores reais da variável dependente e  $\bar{y}$  é a média dos valores reais.

- Calcula-se a Soma dos Quadrados dos Resíduos (SSE), conforme Equação 3.10 (CHEIN, 2019).

- Calcula-se a Soma dos Quadrados da Regressão (SSR), de acordo com Equação 3.9 (CHEIN, 2019).
- Calcula-se  $R^2$ , conforme Equação 3.15 ou 3.16.

$$R^2 = 1 - \frac{SSE}{SST} \quad (3.15)$$

$$R^2 = \frac{SSR}{SST} \quad (3.16)$$

Ambas as fórmulas são equivalentes. A Equação 3.15 primeira expressão subtrai a proporção da variabilidade não explicada pela regressão do total variabilidade, enquanto a Equação 3.16 divide a variabilidade explicada pela regressão pelo total variabilidade.

Um valor de  $R^2$  próximo a -1 ou +1 indica que uma grande proporção da variabilidade na variável dependente é explicada pelas variáveis independentes do modelo. Um valor de  $R^2$  próximo a 0 indica que o modelo não é capaz de explicar a variabilidade na variável dependente.

#### 7. Coeficiente de Determinação $R^2$ ajustado:

O coeficiente de determinação ajustado ( $R^2$  ajustado) é uma métrica estatística que ajusta o  $R^2$  para o número de variáveis independentes em um modelo de regressão. Enquanto o  $R^2$  indica a proporção da variabilidade na variável dependente que é explicada pelo modelo, o  $R^2$  ajustado leva em consideração o número de variáveis independentes incluídas no modelo, penalizando modelos com muitas variáveis independentes irrelevantes.

A fórmula para calcular o  $R^2$  ajustado é dada pela Equação 3.17 (CHEIN, 2019).

$$R^2_{\text{ajustado}} = 1 - \left( \frac{(1 - R^2) \times (n - 1)}{n - k - 1} \right) \quad (3.17)$$

Em que:

- $R^2$ : é o coeficiente de determinação padrão;
- $n$ : é o número total de observações nos dados;
- $k$ : é o número de variáveis independentes no modelo.

Para calcular o  $R^2$  ajustado, deve-se:

- Calcular o  $R^2$ :

Calcula-se o  $R^2$  usando a fórmula padrão mostrada na Equação 3.18, que é a forma expandida das Equações 3.15 ou 3.16.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.18)$$

em que:

- $y_i$  são os valores reais;
- $\hat{y}_i$  são os valores previstos pelo modelo;
- $\bar{y}$  é a média dos valores reais.

- Usar a Fórmula do  $R^2$  ajustado: Substituir o valor calculado de  $R^2$  na fórmula do  $R^2$  ajustado usando o número total de observações  $n$  e o número de variáveis independentes  $k$  no modelo.

– Calcular  $R^2$  ajustado: Realizar as operações matemáticas para calcular o  $R^2$  ajustado. O  $R^2$  ajustado varia de  $-\infty$  a 1, e um valor mais alto indica que uma maior proporção da variabilidade na variável dependente é explicada pelo modelo, levando em consideração o número de variáveis independentes. É uma métrica útil para avaliar a qualidade do modelo, especialmente quando você está comparando modelos com diferentes números de variáveis independentes. Modelos com um  $R^2$  ajustado mais próximo de 1 são geralmente considerados melhores, pois indicam que uma maior porcentagem da variabilidade na variável dependente é explicada pelo modelo.

### 3.3.2 Modelo de Regressão Linear Baeyiana

O segundo modelo estudado na dissertação foi o modelo de Regressão Linear Bayesiana. A inferência bayesiana é baseada na visão subjetiva da probabilidade, ao invés de especificar uma regra para construir um intervalo que conterà o valor verdadeiro de um parâmetro em uma proporção específica de tempo em uma sequência infinita de repetições do mesmo experimento, os modelos bayesianos combinam o conhecimento de um parâmetro disponível antes dos dados serem analisados com informações coletadas durante um experimento para atualizar sua crença sobre o valor do parâmetro após a conclusão do experimento.

Os modelos bayesianos resumem o conhecimento do parâmetro depois de ver os resultados de um experimento usando uma função de densidade de probabilidade. O uso de uma



função de densidade de probabilidade para resumir a incerteza sobre o valor de um parâmetro não significa que os valores de parâmetros desconhecidos sejam aleatórios; significa apenas que o nosso conhecimento do valor de um parâmetro é incerto e que a nossa incerteza sobre este valor pode ser representada utilizando uma função de densidade de probabilidade apropriada (HAMADA *et al.*, 2008).

O modelo de regressão linear bayesiano é uma abordagem estatística que combina os princípios da regressão linear com a teoria bayesiana. Ao contrário dos métodos clássicos, que fornecem estimativas pontuais, a abordagem bayesiana incorpora a incerteza na análise, expressando as estimativas como distribuições de probabilidade. Esse modelo é particularmente útil quando há conhecimento prévio ou informações anteriores disponíveis sobre os parâmetros do modelo.

O Teorema de Bayes fornece um mecanismo formal para atualizar probabilidades. A ideia central por trás da regressão linear bayesiana é aplicar o Teorema de Bayes para atualizar a distribuição de probabilidade dos parâmetros do modelo à medida que mais dados são observados. Inicialmente, são especificadas distribuições probabilísticas a priori para os parâmetros, refletindo o conhecimento prévio ou crenças iniciais. À medida que os dados são incorporados, essas distribuições são atualizadas para fornecer a distribuição a posteriori, que representa a incerteza atualizada sobre os parâmetros do modelo (MORETTIN; BUSSAB, 2010).

O mecanismo para atualizar funções de densidade de probabilidade que representam incerteza sobre o valor de um parâmetro é o Teorema de Bayes. Matematicamente, este teorema pode ser expresso pela Equação 3.19 (HAMADA *et al.*, 2008).

$$p(\theta|y) = \frac{f(y|\theta) \times p(\theta)}{m(y)} \quad (3.19)$$

Em que:

- $p(\theta|y)$ : é a distribuição de probabilidades a posteriori (ou posterior);
- $f(y|\theta)$ : é a distribuição amostral dos dados (ou verossimilhança);
- $p(\theta)$ : é a distribuição de probabilidades a priori (ou prior);
- $m(y)$ : é distribuição marginal dos dados (ou evidência), expressa pela Equação 3.20.

$$m(y) = \int f(y|\theta) \times p(\theta) d\theta \quad (3.20)$$

O elemento a priori  $p(\theta)$  reflete o conhecimento de um parâmetro antes dos dados serem analisados. Uma vez obtidos os dados, a distribuição anterior é atualizada usando as novas

informações. A distribuição de probabilidade atualizada no parâmetro de interesse é chamada de distribuição posterior (ou posterior), porque reflete crenças de probabilidade posteriores ou posteriores à visualização dos dados (HAMADA *et al.*, 2008).

De acordo com o Teorema de Bayes, a distribuição posterior é calculada (até uma constante de proporcionalidade) multiplicando a função de verossimilhança pela distribuição anterior. Dessa forma, a estratégia geral de atualização pode ser expressa de acordo com a Equação 3.21 (HAMADA *et al.*, 2008).

$$posterior \propto verossimilhana \times prior \quad (3.21)$$

Neste contexto, a constante de proporcionalidade absorve o termo  $m(y)$  da Equação 3.21 pois não depende do parâmetro do modelo  $\theta$ . Em termos de funções de densidade de probabilidade, tem-se a Equação 3.22 (HAMADA *et al.*, 2008).

$$p(\theta|y) \propto f(y|\theta) \times p(\theta) \quad (3.22)$$

No contexto da regressão linear bayesiana, os parâmetros incluem os coeficientes da regressão, o termo de interceptação e, opcionalmente, os parâmetros de precisão para as distribuições a priori. As distribuições a priori podem variar, desde distribuições conjugadas (facilitando o cálculo analítico) até distribuições mais flexíveis que refletem a incerteza de maneira mais realista.

O modelo de regressão linear bayesiano oferece várias vantagens, como a capacidade de lidar com pequenas amostras de dados, incorporar informações prévias de especialistas e fornecer estimativas probabilísticas mais ricas. No entanto, a implementação e interpretação podem ser desafiadoras, especialmente para iniciantes na abordagem bayesiana. A diferença fundamental entre um modelo de regressão linear clássico e um modelo de regressão linear bayesiano está na abordagem utilizada para estimar os parâmetros do modelo e lidar com a incerteza associada a essas estimativas. Na regressão linear clássica, os parâmetros do modelo (coeficientes de regressão e intercepto) são estimados por métodos como Mínimos Quadrados Ordinários (MQO). O objetivo é encontrar os valores dos parâmetros que minimizam a soma dos quadrados das diferenças entre as observações reais e as predições do modelo. O modelo de regressão linear clássico assume que os parâmetros são valores fixos e desconhecidos, e não incorpora explicitamente a incerteza associada a essas estimativas.

Na regressão linear bayesiana, os parâmetros são tratados como variáveis aleatórias, e a distribuição de probabilidade dessas variáveis é modelada. A abordagem bayesiana usa o Teorema de Bayes para atualizar as crenças sobre os parâmetros com base nos dados observados. Ao contrário da abordagem clássica, a regressão linear bayesiana incorpora explicitamente a incerteza na estimação dos parâmetros. A distribuição posterior dos parâmetros reflete essa incerteza. Antes de observar os dados, uma distribuição de probabilidade a priori é especificada para os parâmetros, refletindo as crenças iniciais sobre seus valores. Após a observação dos dados, a distribuição a priori é atualizada para a distribuição a posteriori usando o Teorema de Bayes. A inferência bayesiana fornece uma distribuição completa para os parâmetros, o que permite a avaliação direta da incerteza, a geração de intervalos de credibilidade e a tomada de decisões sob incerteza. A abordagem bayesiana é especialmente útil quando há dados limitados, pois permite incorporar informações a priori, o que pode ser de importante na ausência de grandes quantidades de dados. A medida que mais dados são coletados, a distribuição a posteriori é atualizada, permitindo uma adaptação contínua às novas informações.

A base de dados em *.xlsx* foi convertida em *txt* para que fosse possível realizar as simulação. Foi utilizada a linguagem de programação *Python*, que contém bibliotecas próprias para modelagem estatística *Stan*, que trata-se de uma plataforma de última geração para modelagem estatística e computação estatística de alto desempenho, sendo utilizada para modelagem estatística, análise de dados e previsão nas ciências sociais, biológicas e físicas, engenharia e negócios (STAN, 2023).

Para modelagem foram seguidos os seguintes passos:

1. Foram importadas as bibliotecas para modelagem, sendo elas:
  - *pystan*: é uma interface *Python* para o *Stan*, uma linguagem de modelagem estatística e um mecanismo de inferência bayesiana; é frequentemente usada para especificar modelos estatísticos complexos e realizar inferências bayesianas.
  - *pandas*: é amplamente usada para manipulação e análise de dados;
  - *numpy*: é fundamental para computação numérica em *Python*; oferece suporte a *arrays* multidimensionais e funções matemáticas eficientes.
2. Utilizou-se a biblioteca *pandas* para leitura do arquivo em *txt*;
3. Foi nomeada cada coluna do arquivo de acordo com a variável;
4. Foi definido e compilado o modelo em *Stan*:
  - Definiu-se a seção de dados no modelo com o número de amostras e dimensão de

- entradas restritas a valores não negativos;
- Definiu-se a seção de parâmetros no modelo com vetor de parâmetros  $\theta$  e  $\sigma$  restritos a valores não negativos;
  - Definiu-se a seção no modelo com priori normal para  $\theta$  e  $\sigma$  com médias iguais a zero, desvios padrões igual a um e verossimilhança condicional aos dados assumindo uma distribuição normal;
  - O modelo foi compilado e armazenado.
5. Foi criada uma estrutura que dividisse as 343 amostras em pequenos agrupamentos para posteriormente treinar o modelo. No total, foram 43 leilões (conjunto de lotes);
  6. Foi definida uma função de predição que aceitasse os parâmetros das entradas e de  $\theta$ . A operação é uma multiplicação de matrizes, onde  $X$  é a matriz de entrada e  $\theta$  é o vetor de parâmetros. O resultado,  $y$ , é a predição.
  7. Foram avaliados o desempenho do modelo estatístico em comparação aos valores reais
    - Atribuiu-se os valores verdadeiros a uma variável;
    - Atribuíram-se os valores estimados e os limites inferiores e superiores às suas respectivas variáveis;
    - Calculou-se a probabilidade de cobertura positiva, ou seja, a proporção de intervalos de confiança que incluem os verdadeiros parâmetros. A probabilidade de cobertura é uma métrica que indica a proporção de intervalos de confiança que contêm o verdadeiro valor da variável de interesse. É uma medida de quão bem os intervalos de confiança calculados pelo modelo correspondem aos resultados reais; a proporção de intervalos de confiança que incluem os verdadeiros parâmetros. Pode-se pensar nesta quantidade como a chance de o intervalo construído por tal procedimento conter o parâmetro de interesse (STATISTICS, 2023);
    - Calculou-se a largura média dos intervalos de confiança; A largura média dos intervalos de confiança é a média das larguras de todos os intervalos de confiança calculados pelo modelo. Isso fornece uma medida da precisão dos intervalos - quanto menor a largura média, mais precisa é a estimativa.
    - Calculou-se o erro quadrático médio global (RMSE); o RMSE é uma medida da diferença entre os valores previstos e os valores reais. Ele penaliza erros maiores de forma mais significativa do que erros menores. Quanto menor o RMSE, melhor é o ajuste do modelo aos dados.

- Calculou-se o RMSE apenas para os casos em que os intervalos de confiança eram positivos. Esta métrica foca especificamente nos casos onde as previsões e os valores reais são positivos. Pode ser útil se houver uma assimetria ou importância especial associada aos casos positivos.
8. Foi realizado o ajuste do modelo para leilão (conjunto de lotes) em um conjunto de dados:
- É realizado um *loop* sobre cada leilão;
  - Para cada conjuntos de dados (leilão), foram separados os dados de treino e teste e normalizadas as características;
  - Ajustou-se o modelo estatístico bayesiano usando o modelo Stan com os dados de treino;
  - Obteve-se os valores estimados do modelo;
  - Realizou-se previsões usando diferentes percentis dos parâmetros (50%, 5%, 95%);
  - Avaliou-se o desempenho do modelo.
9. Por fim, calculou-se a média das probabilidades da proporção de intervalos de confiança que contêm o valor verdadeiro do parâmetro que está sendo estimado.

## 4 RESULTADOS

Neste capítulo serão apresentados os resultados obtidos nos dois métodos utilizados para modelagem.

### 4.1 Resultados do Modelo de Regressão Linear Múltiplo

Com o auxílio do *Matlab* foram calculados média e desvio padrão das variáveis, que é uma etapa importante para a compreensão da variabilidade dos dados utilizados no modelo de regressão linear múltipla. Na Tabela 5 pode-se observar os resultados de média e variância calculados

Tabela 5 – Média e desvio padrão

Variável	Média	Desvio Padrão
num_comp	7,49	3,68
emp	3,66	4,23
ext	276,95	391,45
prz	36,52	15,28
nrt	0,14	0,35
ne	0,32	0,48
sul	0,17	0,38
se_co	0,44	0,50
estat	0,34	0,47
estrang	0,75	0,43
cons	0,82	0,38
dif	20,74	36,33

Fonte: Elaborado pelo autor.

Após esta primeira etapa, foi realizado um escalonamento nos dados, de modo que a média fosse igual a zero e o desvio padrão igual a um, de modo assegurar que as variáveis tenham a mesma escala e evitar que alguma delas exerça maior influência no modelo de regressão. Após o escalonamento, foi verificada a correlação entre as variáveis independentes e a saída, conforme Tabela 6.

Para demonstrar melhor os resultados de correlação, foi plotado o mapa de calor conforme mostrado na Figura 11.

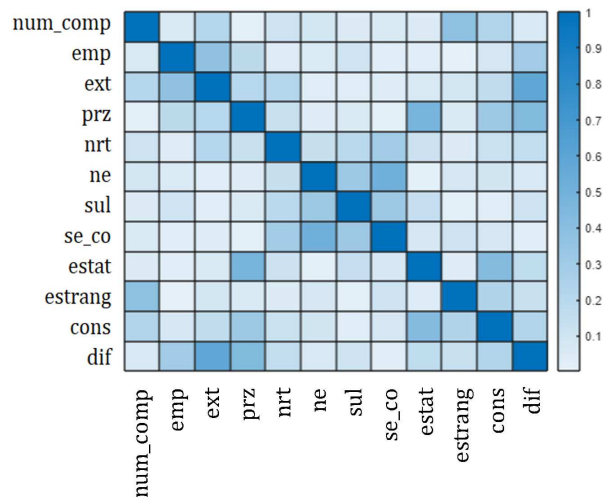
Analisando a Tabela 6 e a Figura 11, pode-se observar que o maior índice de correlação foi entre a variável *ext* e a variável dependente *dif*, ou seja, há uma correlação mais forte entre essas variáveis. Os resultados de correlação obtidos com o modelo de regressão linear múltipla mostraram uma baixa correlação entre as variáveis, ou seja, os valores encontrados foram mais próximos de 0 e mais distantes de + ou -1.

Tabela 6 – Correlação entre as variáveis

	num_comp	emp	ext	prz	nrt	ne	sul	se_co	estat	estrang	cons	dif
num_comp	1,000	-0,062	-0,212	0,016	-0,111	0,081	0,045	-0,055	0,044	0,377	0,224	0,059
emp	-0,062	1,000	0,376	0,178	0,039	-0,047	0,094	-0,024	0,026	0,002	0,074	0,302
ext	-0,212	0,376	1,000	0,202	0,211	-0,031	0,026	0,041	-0,061	0,081	0,159	0,579
prz	0,016	0,178	0,202	1,000	0,130	0,034	-0,057	0,012	-0,475	-0,056	0,313	0,432
nrt	-0,111	0,039	0,211	0,130	1,000	-0,133	-0,186	-0,290	-0,113	-0,044	0,122	0,155
ne	0,081	-0,047	-0,031	0,034	-0,133	1,000	-0,316	-0,509	0,006	-0,074	-0,089	-0,060
sul	0,045	0,094	0,026	-0,057	-0,186	-0,316	1,000	-0,316	0,138	0,012	-0,027	0,102
se_co	-0,055	-0,024	0,041	0,012	-0,290	-0,509	-0,316	1,000	-0,068	0,109	0,075	0,019
estat	0,044	0,026	-0,061	-0,475	-0,113	0,006	0,138	-0,068	1,000	0,038	-0,421	-0,166
estrang	0,377	0,002	0,081	-0,056	-0,044	-0,074	0,012	0,109	0,038	1,000	0,232	0,131
cons	0,224	0,074	0,159	0,313	0,122	-0,089	-0,027	0,075	-0,421	0,232	1,000	0,219
dif	0,059	0,302	0,579	0,432	0,155	-0,060	0,102	0,019	-0,166	0,131	0,219	1,000

Fonte: Elaborado pelo autor.

Figura 11 – Mapa de calor da correlação entre as variáveis



Fonte: Elaborado pelo autor.

Os resultados dos parâmetros do modelo, valores de erro padrão dos coeficientes, de estatística-t e de valor p, estão apresentados na Tabela 7.

Tabela 7 – Resultados obtidos - Regressão Linear Múltipla

Variável	Coefficiente	Er. pad.	Estat.-t	Valor-p
num_comp	0,1549	0,0462	33,548	8,865e-04
emp	0,0562	0,0436	12,898	0,198
ext	0,4973	0,0479	10,371	5,230e-18
prz	0,3172	0,0473	67,089	8,510e-08
nrt	0,1002	0,0636	1,577	0,116
ne	0,0609	0,0798	0,764	0,445
sul	0,1687	0,0714	23,625	0,019
se_co	0,1133	0,0850	13,332	0,183
estat	-0,0099	0,0490	-0,202	0,840
estrang	0,0498	0,0446	11,166	0,265
cons	-0,0240	0,0476	-0,505	0,614

Fonte: Elaborado pelo autor.

O modelo de regressão linear múltipla apresentou um coeficiente de determinação  $R^2$  ajustado igual a 0,468, o que significa que aproximadamente 46,8% da variabilidade da diferença entre a RAP proposta pela ANEEL e a RAP do proponente vencedor é explicada pelas variáveis independentes utilizadas no modelo, controlando o efeito do tamanho da amostra e do número de variáveis no modelo. Isso significa que as variáveis independentes incluídas no modelo de regressão linear múltipla, como o número de competidores, o número de empreendimentos, a extensão da linha de transmissão e o prazo de execução da obra, foram relevantes para entender e prever a diferença entre os valores de RAP propostos pela ANEEL e os valores de RAP dos proponentes vencedores nos leilões.

É importante observar os valores do erro padrão dos coeficientes na Tabela 7 para cada uma das variáveis independentes. Esses valores indicam a variabilidade das estimativas dos coeficientes em relação à média e fornecem uma ideia da precisão das estimativas. Quanto menor o valor do erro padrão, mais preciso é o coeficiente estimado e menor é a incerteza associada a ele.

Como mostrado anteriormente, um valor-p baixo ( $< 0,05$ ) indica que a hipótese nula pode ser rejeitada, o que sugere que o preditor com um valor-p baixo provavelmente terá um impacto significativo no modelo, já que as alterações nesse preditor estão relacionadas a mudanças na variável resposta, dessa forma, observa-se que apenas as variáveis num\_comp, ext, prz, e sul estão associada a mudanças significativas na diferença entre os valores de RAP. Isso indica que as demais variáveis não contribuem de forma relevante para explicar a variabilidade dos dados.

De posse dos resultados, verificou-se que o modelo inicialmente escolhido para modelagem não apresentou resultados satisfatórios. Foram realizados testes excluindo algumas das variáveis menos significativas, entretanto, o teve pouca alteração nos resultados. Dessa forma, consultou-se a literatura sobre outros modelos e optou-se por utilizar um modelo de regressão Linear Bayesiano.

## **4.2 Resultados do Modelo de Regressão Linear Bayesiana**

Após realizada a modelagem conforme descrito na seção 3.3, obteve-se os resultados de métricas de desempenho que são apresentados na Tabela 8.

Os resultados fornecem uma avaliação do desempenho do modelo estatístico bayesiano para cada leilão. Analisando os resultados das métricas obtidos tem-se que:



Tabela 8 – Resultados obtidos - Regressão Linear Bayesiana

Núm. Leilão	Probabilidade de Cobertura	Largura Média	RMSE	RMSE Positivos	Média da Estimativa
0	0,857	26,911	9,810	7,129	20,518
1	0,900	26,247	12,122	10,112	20,867
2	0,500	29,711	33,257	7,946	30,095
3	1,000	33,760	12,028	12,028	29,693
4	1,000	30,155	11,406	11,406	18,299
5	1,000	27,790	11,444	11,444	22,529
6	1,000	28,769	11,140	11,140	21,779
7	1,000	30,533	18,522	18,522	21,566
8	1,000	21,877	8,106	8,106	10,797
9	1,000	33,950	73,155	73,155	51,458
10	1,000	83,264	60,773	60,773	68,295
11	1,000	26,788	17,819	17,819	18,566
12	1,000	21,617	9,110	9,110	13,678
13	1,000	22,979	10,392	10,392	12,780
14	1,000	21,630	7,781	7,781	14,271
15	1,000	25,576	8,287	8,287	13,226
16	0,667	21,981	13,882	7,437	14,324
17	1,000	24,121	15,422	15,422	16,890
18	1,000	22,179	15,504	15,504	17,109
19	0,800	24,008	20,515	7,749	22,510
20	1,000	15,244	8,112	8,112	10,178
21	1,000	19,158	17,745	17,745	13,728
22	1,000	31,656	31,975	31,975	37,653
23	1,000	33,535	26,278	26,278	33,376
24	1,000	29,352	16,193	16,193	17,284
25	1,000	28,772	26,427	26,427	24,403
26	0,000	60,449	175,009	nan	91,391
27	1,000	28,558	18,827	18,827	17,758
28	1,000	29,504	16,816	16,816	21,514
29	1,000	35,828	34,642	34,642	40,749
30	1,000	27,149	34,828	34,828	29,312
31	1,000	32,663	23,976	23,976	22,907
32	0,000	75,925	117,194	nan	114,566
33	1,000	25,320	26,709	26,709	22,950
34	1,000	32,445	18,092	18,092	28,647
35	0,774	33,960	35,118	10,860	22,142
36	0,636	39,176	41,986	8,558	32,027
37	0,850	29,954	26,522	10,183	17,529
38	0,562	38,140	59,733	6,124	32,437
39	0,750	32,902	22,983	12,071	21,913
40	0,500	35,922	44,398	13,988	22,067
41	1,000	31,158	8,026	8,026	17,744
42	0,600	34,236	35,709	7,634	21,327

Fonte: Elaborado pelo autor.

- Probabilidade de Cobertura (*Coverage Probability*): representam a proporção de intervalos de confiança que realmente contêm o valor verdadeiro; valores próximos a 1 indicam uma boa cobertura. Os leilões 0, 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 23, 24, 25, 27, 28, 29, 30, 31, 33, 34, 35, 37, 39 e 41 mostram uma cobertura positiva

próximas ou iguais a 1.

- Largura Média (*Average Width*): reflete a precisão das estimativas; quanto menor a largura média, melhor. Dentre os resultados obtidos, pode-se citar leilões 8, 12, 13, 14, 16, 17, 18, 19, 20, 21 e 33 apresentaram larguras médias menores em relação aos resultados dos outros leilões.
- RMSE (*Root Mean Squared Error*): calcula o erro quadrático médio global; quanto menor o valor obtido, melhor o resultado. Os leilões 0, 4, 6, 8, 12, 13, 14, 15, 20 e 41 têm valores mais baixos em relação aos valores de RMSE dos outros leilões;
- RMSE Positivo: calcula o RMSE apenas para os casos em que os intervalos de confiança são positivos; quanto menor o valor obtido, melhor o resultado. Os leilões 0, 4, 6, 8, 12, 13, 14, 15, 20 e 41 têm valores mais baixos em relação aos valores de RMSE positivo dos outros leilões. Para os valores "nan" pode indicar que não há previsões positivas ou que não foram observados valores verdadeiros positivos naquele conjunto de dados específico.
- Média dos Leilões (*AverageLeilões*): média das estimativas dos resultados dos leilões.

De modo geral, o modelo geralmente apresenta valores altos (próximo a 1) de probabilidade de cobertura, indicando que os intervalos de confiança são precisos. Dessa forma, foi calculada a média das probabilidades de cobertura dos leilões é de 0,8697 ou 86,97%, o que significa que o modelo explica aproximadamente 87% da variabilidade da variável dependente *RAP\_ed - RAP\_prop*.

Quanto aos parâmetros do modelo, obteve-se os resultados mostrados na Tabela 9.

Tabela 9 – Saída do modelo no Stan

Parâmetro	Mean	SE_Mean	SD	2,5%	25%	50%	75%	97,5%	n_eff	Rhat
$\theta$ num_comp	0,04	8,60e-04	0,03	1,50e-03	0,02	0,04	0,06	0,12	1390	1
$\theta$ emp	0,46	1,10e-03	0,05	0,37	0,43	0,46	0,49	0,55	1664	1
$\theta$ ext	0,06	8,00e-04	0,03	0,01	0,04	0,06	0,08	0,11	1103	1
$\theta$ prz	0,05	8,50e-04	0,03	2,60e-03	0,02	0,04	0,07	0,12	1489	1
$\theta$ nrt	0,02	3,20e-04	0,02	6,30e-04	5,30e-03	0,01	0,02	0,06	2380	1
$\theta$ ne	0,06	8,20e-04	0,04	6,60e-03	0,03	0,06	0,08	0,14	1878	1
$\theta$ sul	0,02	3,60e-04	0,02	8,00e-04	6,00e-03	0,01	0,03	0,06	2239	1
$\theta$ se_co	0,02	2,90e-04	0,01	5,30e-04	5,60e-03	0,01	0,02	0,05	2281	1
$\theta$ estat	8,50e-03	1,70e-04	8,20e-03	1,70e-04	2,60e-03	6,10e-03	0,01	0,03	2399	1
$\theta$ estrang	0,01	2,60e-04	0,01	3,30e-04	4,20e-03	0,01	0,02	0,05	2388	1
$\theta$ cons	0,01	2,60e-04	0,01	4,00e-04	4,10e-03	9,50e-03	0,02	0,04	1954	1
sigma	0,8	7,00e-04	0,03	0,74	0,78	0,8	0,82	0,86	1845	1
lp	-135,6	0,1	2,66	-141,8	-137,3	-135,5	-133,7	-131,4	646	1

Fonte: Elaborado pelo autor.

A Tabela 9 retorna a saída após ajuste de um modelo no Stan. As estatísticas obtidas

foram:

- Mean: média dos parâmetros  $\theta$  de cada variável;
- SE\_Mean: média do erro padrão de cada variável;
- SD: o desvio padrão dos valores amostrados;
- Percentis de distribuição posterior: são calculados os percentis 2,5%, 25%, 50%, 75%, e 97,5%.
- n\_eff: uma medida aproximada do tamanho efetivo da amostra para o parâmetro. Um valor alto de n\_eff é desejado, pois sugere uma amostra efetiva grande, o que geralmente indica que as amostras são informativas para a inferência; valores baixos de n\_eff indicam alta autocorrelação entre as amostras, o que pode diminuir a eficácia da amostra para representar a distribuição verdadeira.
- Rhat: o fator de redução de escala potencial, usado para verificar a convergência; um valor de Rhat próximo de 1 indica convergência.

A variável  $lp$  (*log-probability*) refere-se à log-verossimilhança posterior, que é a probabilidade de observar os dados (dado o modelo e os parâmetros) expressa no espaço logarítmico.

Analisando os percentis obtidos na Tabela 9, tem-se que:

- Para os percentis 2,5% e 97,5% (intervalo de 95%): os intervalos de confiança para todos os parâmetros são relativamente estreitos, indicando uma precisão razoável nas estimativas;
- Para os percentis 25%, 50% (mediana) e 75%: Os percentis 25%, 50%, e 75% também não mostram uma grande dispersão; os valores estão relativamente próximos, sugerindo consistência nas estimativas.

Quanto aos valores de Rhat (diagnóstico de convergência do modelo), todos os valores de Rhat são próximos de 1, o que é um bom sinal, o que sugere que houve convergência no modelo.

Os valores de  $\theta$  representam os coeficientes de regressão do modelo, dessa forma, as variáveis mais significativas identificadas pelo modelo foram as variáveis número de empreendimentos, extensão da linha de transmissão, empreendimentos realizados no subsistema nordeste, prazo para execução do empreendimento e número de competidores.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

Neste capítulo serão apresentadas as considerações finais com as conclusões da pesquisa, as sugestões de trabalhos futuros e as publicações realizadas durante o período de mestrado.

### 5.1 Considerações Finais

Nesta dissertação foram utilizados dois modelos de regressão para analisar o impacto de variáveis que pudessem influenciar no valor final da diferença entre a RAP do edital e a RAP proposta em leilões de transmissão de energia: um linear e um bayesiano. Foi analisado um conjunto de 343 observações referentes aos resultados dos leilões de transmissão de energia elétrica, considerando-se 11 variáveis independentes: número de competidores do leilão por lote (*num\_comp*), número de empreendimentos por lote (*emp*), extensão das linhas de transmissão (*ext*) e prazo para execução do empreendimento (*prz*), variáveis *dummy* para empreendimentos na subsistema norte (*nrt*), nordeste (*ne*), sul (*sul*) e/ou sudeste/centro-oeste (*se\_co*) e variáveis *dummy* para participação de empresas estatais (*estat*), estrangeiras (*estrang*) e/ou participação em consórcio (*cons*).

Na primeira etapa utilizou-se um modelo de regressão linear múltiplo, onde observou-se que as variáveis levantadas conseguem explicar aproximadamente 46,8% da variabilidade da variável dependente  $RAP_{ed} - RAP_{prop}$ . Pôde-se verificar que muitas das variáveis apresentaram uma fraca correlação entre si, onde a maior correlação encontrada foi entre as variáveis *ext* (extensão da linha de transmissão) e *dif* (variável dependente - diferença entre RAP do edital e RAP proposto).

Com o resultado parcial obtido, foi possível verificar que uma parcela considerável da variabilidade que não foi explicada pelo modelo, que poderia ser atribuído a outras variáveis ou fatores que não foram considerados, bem como à presença de componentes de incerteza e aleatoriedade nos dados. Portanto, é necessário ter cautela ao interpretar os resultados e reconhecer que há outros elementos que podem influenciar a diferença entre os valores de RAP.

O modelo de regressão múltipla é uma ótima ferramenta estatística muito utilizada em econometria, mas que apresenta algumas limitações que podem afetar as conclusões das análises, entre elas a autocorrelação dos resíduos, heterocedasticidade e multicolinearidade. Após analisar os resultados obtidos nos trabalhos anteriores e verificar as limitações do modelo e

que o resultado encontrado justifica apenas cerca de metade dos resultados dos leilões no período em questão, optou-se por utilizar outro modelo para análise a fim de se obter melhores resultados e sanar o problema com as limitações do modelo atual: o modelo de Regressão Linear Bayesiana.

Com o novo modelo utilizado os resultados tiveram uma melhoria considerável. Foram realizados treinamentos em 43 conjuntos lotes e eles resultaram uma média de cobertura positiva de 86,97%, ou seja, as variáveis neste modelo explicam aproximadamente 87% da variabilidade da variável dependente *RAP<sub>ed</sub> - RAP<sub>prop</sub>*.

Para se propor um modelo de previsão do resultado da diferença entre a Receita Anual Permitida do edital e a proposta pelo competidor foram utilizados dois métodos de regressão: um linear, também chamado de frequentista, e um bayesiano. O modelo linear apresentou algumas limitações e obteve um resultado abaixo do esperado, o que motivou a utilização de outro método.

Para além dos resultados numéricos obtidos nesta dissertação, destaca-se a importância e contribuição do estudo realizada em Teoria dos Leilões e estudo de leilões de transmissão de energia elétrica no Brasil e bem como a criação de um faustoso banco de dados com informações históricas que podem ser amplamente utilizadas para outros estudos, como a análise com outros modelos e obtenção de resultados cada vez mais próximos com o real.

## 5.2 Trabalhos Futuros

Como proposta de continuidade deste trabalho sugere-se:

- a utilização de técnicas de Inteligência Artificial, como máquinas de aprendizado, redes neurais, entre outros, para melhorar a predição do lance vencedor;
- o estudo de outras variáveis, principalmente financeiras, como o Risco Brasil, que possam justificar o lance vencedor no leilão de transmissão de energia;
- a realização de testes com leilões que ainda vão acontecer para validação do modelo;
- a aplicação da mesma metodologia - Regressão Linear Bayesiana para previsão dos resultados de leilão de geração de energia no ambiente regulado (ACR);

## 5.3 Publicações

Como resultado desta dissertação realizou-se a publicação do artigo intitulado *Análise de Fatores que Impactam o Valor do Lance Vencedor nos Leilões de Transmissão de Energia Elétrica* (OLIVEIRA *et al.*, 2023) em 2023 no Simpósio Brasileiro de Sistemas Elétricos (SBSE).

Neste trabalho foram utilizados como variável dependente a diferença entre a RAP do edital pela RAP proposta e como variáveis independentes o número de competidores, o número de empreendimentos, a extensão das linhas de transmissão e o prazo para construção do empreendimento em um modelo de regressão linear múltiplo. Neste artigo, as variáveis analisadas justificavam em 45,5% as o resultado dos leilões no período analisado.

Além disso, participou-se também da escrita do artigo *O hidrogênio verde no Brasil e as necessidades de um arcabouço regulatório* (SILVA *et al.*, 2023) com publicação em 2023 no XXVII Seminário Nacional de Produção e Transmissão de Energia Elétrica. Neste trabalho apresentou-se uma discussão sobre as políticas públicas e as decisões regulatórias necessárias para alcançar um equilíbrio entre a eficiência do sistema elétrico do Brasil, a manutenção de tarifas acessíveis e o progresso no processo de transição energética.

## REFERÊNCIAS

- ANEEL. **Base ANEEL**. 2023. Disponível em: <https://www.gov.br/aneel/pt-br/centrais-de-conteudos/relatorios-e-indicadores/leiloes>. Acesso em: 16 ago. 2023.
- ANEEL. **Editais de Transmissão ANEEL**. 2023. Disponível em: [https://www2.aneel.gov.br/aplicacoes\\_liferay/editais\\_transmissao/edital\\_transmissao.cfm](https://www2.aneel.gov.br/aplicacoes_liferay/editais_transmissao/edital_transmissao.cfm). Acesso em: 16 ago. 2023.
- ANEEL. **EDITAL DO LEILÃO Nº 1/2023-ANEEL**. 2023. Disponível em: [https://www2.aneel.gov.br/aplicacoes/editais\\_transmissao/documentos/Edital\\_Leil%C3%A3o\\_1\\_2023\\_ANEEL\\_publica%C3%A7%C3%A3o.pdf](https://www2.aneel.gov.br/aplicacoes/editais_transmissao/documentos/Edital_Leil%C3%A3o_1_2023_ANEEL_publica%C3%A7%C3%A3o.pdf). Acesso em: 16 ago. 2023.
- B3. **Resumo leilões**. 2023. Disponível em: <https://bvmf.bmfbovespa.com.br/consulta-leiloes/ResumoAcompanhamentoLeilao.aspx?idioma=pt-br>. Acesso em: 16 ago. 2023.
- BEN. **Balanco Energético Nacional**. 2010. Disponível em: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/balanco-energetico-nacional-2010>. Acesso em: 16 ago. 2023.
- BEN. **Balanco Energético Nacional**. 2023. Disponível em: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/balanco-energetico-nacional-2023>. Acesso em: 16 ago. 2023.
- CASTRO, N. d.; MARTINI, S. Como explicar os resultados dos leilões de transmissão? GESEL, 2021.
- CAZZARO, P. M. **Análise da dinâmica dos investidores nos leilões de transmissão de energia elétrica no Brasil entre 1999 e 2017**. 2017. 140 f. Dissertação (Mestrado em Ciências) – Programa de Pós-Graduação do Instituto de Energia e Ambiente, Universidade de São Paulo, São Paulo - SP, 2017.
- CEZARIO, A.; RAMOS, F.; SOUZA, H. Fatores influenciadores nos lances vencedores dos leilões de transmissão. XLI SBPO, 2009, 2009.
- CHEIN, F. **Introdução aos Modelos de Regressão Linear**. [S. l.]: ENAP, 2019. v. 2019.
- COLELL, A.; WHINSTON, M. D.; GREEN, J. R. **Microeconomic Theory**. [S. l.]: Oxford University Press, USA, 1995. v. 1995.
- EPE. **Estudos do Plano Decenal de Expansão de Energia 2032 - Transmissão de Energia**. 2023. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-689/topico-640/Caderno%20de%20Transmiss%C3%A3o%20de%20Energia%20-%20PDE%202032.pdf>. Acesso em: 16 ago. 2023.
- EPE. **Plano Decenal de Expansão de Energia 2032 - Demanda de Eletricidade**. 2023. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-689/topico-640/Caderno%20de%20Demanda%20de%20Eletricidade%20-%20PDE%202032.pdf>. Acesso em: 16 ago. 2023.
- FIANI, R. **Teoria dos Jogos**. [S. l.]: Elsevier, 2009. v. 2018.

FILHO, N. S. C. **Uma análise comparativa da aplicação da teoria dos jogos aos vários tipos de leilões em mercados elétricos.** 2017. 90 f. Dissertação (Mestrado em Engenharia de Eletricidade) – Programa de Pós-Graduação em Engenharia de Eletricidade/CCET, Universidade Federal do Maranhão, São Luiz - MA, 2017.

FRACASSO, B. **LEILÕES DE TRANSMISSÃO DE ENERGIA ELÉTRICA: DETERMINANTES DOS DESÁGIOS NOS LEILÕES DE 2011 A 2018.** 30 f. Monografia (Monografia) – Escola Nacional de Administração Pública, Brasília, 2019.

HAMADA, M. S.; WILSON, A. G.; REESE, C. S.; MARTZ, H. F. **Bayesian Reliability.** [S. l.]: Springer, 2008. v. 2008.

KRISHNA, V. **Auction Theory.** [S. l.]: Academic Press, 2002. v. 2002.

MENEZES, F. M.; MONTEIRO, P. K. **An Introduction to Auction Theory.** [S. l.]: Oxford University Press, 2008. v. 2008.

MOCHÓN, A.; SáEZ, Y. **Understanding Auctions.** [S. l.]: Springer, 2015. v. 2015.

MONTGOMERY, D.; RUNGER, G. **Estatística aplicada e probabilidade para engenheiros.** [S. l.]: LTC, 2009. v. 2021.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística Básica.** [S. l.]: Editora Saraiva, 2010. v. 2010.

NASCIMENTO, R. L. **ANÁLISE DOS FATORES DE INFLUÊNCIA NAS PROPOSTAS OFERTADAS NOS LEILÕES DE TRANSMISSÃO DE ENERGIA ELÉTRICA.** 2012. 46 f. Dissertação (Mestrado em Economia do Setor Público) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação / Departamento de Economia, Brasília, 2012.

OLIVEIRA, A.; SOUZA, F.; PONTES, R. **Análise de fatores que impactam o valor do lance vencedor nos leilões de transmissão de energia elétrica.** SBSE,2023, 2023.

ONS. **o que é o SIN.** 2023. Disponível em: <https://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>. Acesso em: 16 ago. 2023.

ONS. **Sistema Interligado Nacional.** 2023. Disponível em: <https://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>. Acesso em: 16 ago. 2023.

ONS. **Sistemas Isolados.** 2023. Disponível em: <https://www.ons.org.br/paginas/sobre-o-sin/sistemas-isolados>. Acesso em: 16 ago. 2023.

PAULA, G. A. **Modelos de Regressão com Apoio Computacional.** [S. l.]: Instituto de Matemática e Estatística, 2004. v. 2004.

RASO, A. A. D. **LEILÕES DE TRANSMISSÃO DE ENERGIA: UMA ANÁLISE DA CONCORRÊNCIA.** 58 f. Monografia (Monografia) – Ciências Econômicas, Universidade Federal da Bahia, Salvador, 2013.

SARTINI, B.; GARBUGIO, G.; BORTOLOSSI, H.; SANTOS, P.; BARRETO, L. **Uma Introdução a teoria dos Jogos.** [S. l.]: II Biental da SBM, 2004. v. 2004.

SILVA, F.; SILVA, H.; PONTES, R.; OLIVEIRA, A.; CORREIA, F.; AGUIAR, V. **O hidrogênio verde no brasil e as necessidades de um arcabouço regulatório.** SNPTEE,2023, 2023.



STAN. **Stan**. 2023. Disponível em: <https://mc-stan.org/>. Acesso em: 16 ago. 2023.

STATISTICS, I. for the Advancement of University Learning Department of. **Confidence Intervals**. 2023. Disponível em: <https://www.stats.ox.ac.uk/pub/bdr/IAUL/Course1Notes5.pdf>. Acesso em: 16 ago. 2023.

## APÊNDICE A – MODELAGEM - REGRESSÃO LINEAR MÚLTIPLA

Segue na Figura 12 o código utilizado para os cálculos dos resultados obtidos na dissertação.

Figura 12 – Código no Matlab utilizado para modelagem

```
x = load("22.10.v1.txt"); %carregar o banco de dados
md=mean(x); %calcular média
dp=std(x); %calcular desvio padrão
x=mapstd(x'); x=x'; %escalonar os dados
num_comp=x(:,1); %nomear as colunas das variáveis
emp=x(:,2);
ext=x(:,3);
prz=x(:,4);
nrt=x(:,5);
ne=x(:,6);
sul=x(:,7);
se_co=x(:,8);
estat=x(:,9);
estrang=x(:,10);
cons=x(:,11);
dif=x(:,12);
X=[num_comp, emp, ext, prz, nrt, ne, sul, se_co, estat, estrang, cons];
Xn=[num_comp, emp, ext, prz, nrt, ne, sul, se_co, estat, estrang, cons, dif];
Y=fitlm(X,x(:,end)); %calcular estatísticas do modelo
corrcoef(Xn); %verificar correlação entre as variáveis
```

Fonte: Elaborado pelo autor.

## APÊNDICE B – MODELAGEM - REGRESSÃO LINEAR BAYESIANA

Segue abaixo o código utilizado para modelo de regressão linear bayesiana.

```
1
2 !pip install git+https://github.com/stan-dev/pystan2.
   git@master
3
4 import pystan
5 import pandas as pd
6 import numpy as np
7 import matplotlib.pyplot as plt
8
9 data.keys()
10
11 leiloes_code = ""
12 data {
13     int<lower=0> N; // numero de amostras
14     int<lower=0> D; // dimensao das entradas
15     matrix[N,D] X; // entradas
16     vector[N] y; // saida
17 }
18 parameters {
19     vector<lower=0>[D] theta;
20     real<lower=0> sigma;
21 }
22 transformed parameters {
23
24 }
25 model {
26
27     theta ~ normal(0,1);
```

```
28     sigma ~ normal(0,1);
29
30     y ~ normal(X*theta, sigma);
31 }
32 """
33
34 sm = pystan.StanModel(model_code=leiloes_code)
35
36 tmp = {
37     'Start': [1, 8, 18, 20, 27, 33, 40, 47, 59, 65, 72, 75,
38              86, 94, 103, 106, 113, 116, 128, 136, 141, 144,
39              149, 156, 162, 167, 177, 178, 181, 189, 193, 195,
40              199, 200, 214, 234, 265, 276, 296, 312, 324, 334,
41              339],
42     'End': [7, 17, 19, 26, 32, 39, 46, 58, 64, 71, 74, 85,
43            93, 102, 105, 112, 115, 127, 135, 140, 143, 148,
44            155, 161, 166, 176, 177, 180, 188, 192, 194, 198,
45            199, 213, 233, 264, 275, 295, 311, 323, 333, 338,
46            343]
47 }
48 data_start_end_leiloes = pd.DataFrame(tmp)
49 data_start_end_leiloes.head(2)
50
51 def predict(X,theta):
52     y = X@theta
53
54     return y
55
56 import numpy as np
57
58 def performance(y_test,yest50,yest5,yest95 ):
59     # True parameter values
```

```
52 true_params = y_test
53
54 # Estimated parameter values and confidence intervals
55 estimated_params = yest50
56 lower_bounds = yest5
57 upper_bounds = yest95
58
59 # Positive coverage
60 coverage_positive = np.sum(upper_bounds >= true_params)/
    len(true_params)
61
62 # Width of Confidence Intervals
63 widths = np.array(upper_bounds) - np.array(lower_bounds)
64 average_width = np.mean(widths)
65
66 # Root Mean Squared Error (MSE)
67 squared_errors = (np.array(estimated_params) - np.array(
    true_params)) ** 2
68 rmse = np.sqrt(np.mean(squared_errors))
69
70 # Root Mean Squared Error (RMSE)
71 ind_positive = upper_bounds >= true_params
72 positive_squared_errors = (np.array(estimated_params[
    ind_positive]) - np.array(true_params[ind_positive]))
    ** 2
73 positive_rmse = np.sqrt(np.mean(squared_errors[
    ind_positive]))
74
75 # Print the metrics
76 print(f"Coverage Probability: {coverage_positive:.3f}")
77 print(f"Average Width: {average_width:.3f}")
78 print(f"RMSE: {rmse:.3f}")
```

```
79 print(f"Positive RMSE: {positive_rmse:.3f}")
80 print(f"Average Leiloes: {np.mean(yest50):.3f}")
81
82 return coverage_positive, average_width, rmse,
      positive_rmse, np.mean(yest50)
83
84 coverage_positive_i = []
85 average_width_i = []
86 rmse_i = []
87 positive_rmse_i = []
88 mean_lance_i = []
89
90 for i in range(len(data_start_end_leiloes)):
91
92     n_leilao = i
93     start_leilao = data_start_end_leiloes.loc[
94         data_start_end_leiloes.index==n_leilao]['Start'].
95         values[0]
96     end_leilao = data_start_end_leiloes.loc[
97         data_start_end_leiloes.index==n_leilao]['End'].values
98         [0]
99     ind_test = np.arange(start_leilao-1,end_leilao)
100     ind_train = np.arange(0,343)
101     ind_train = [x for x in ind_train if x not in ind_test]
102
103     inputs = ['emp', 'ext', 'prz', 'nrt', 'ne', 'sul', 'se_co
104         ', 'num_comp', 'estat', 'estrang', 'cons']
105
106     X=data[inputs].iloc[ind_train]
107     labels = X.keys()
108     x_std = X.std()
109     X = X.div(x_std).values
```

```
105 D=X.shape[1]
106 y=data[['dif']].iloc[ind_train].values[:,0]
107 y_std = np.std(y)
108 y=y/y_std
109
110 X_test = data[inputs].iloc[ind_test]
111 y_test = data[['dif']].iloc[ind_test].values[:,0]
112 X_test = X_test.div(x_std).values
113 y_test = y_test/y_std
114
115 N = X.shape[0]
116
117 data_leiloes = {'N': N,
118                'D': D,
119                'X': X,
120                'y': y}
121
122 fit = sm.sampling(data=data_leiloes, iter=1000, chains=4)
123
124 theta = fit['theta']
125 theta50 = np.percentile(theta,50,axis=0)
126 theta99 = np.percentile(theta,99,axis=0)
127 theta1 = np.percentile(theta,1,axis=0)
128 theta95 = np.percentile(theta,95,axis=0)
129 theta5 = np.percentile(theta,5,axis=0)
130
131 yest50 = predict(X_test,theta50)
132 yest5 = predict(X_test,theta5)
133 yest95 = predict(X_test,theta95)
134 yest1 = predict(X_test,theta1)
135 yest99 = predict(X_test,theta99)
136
```

```
137 print(' - - - - - ')
138 print(f'Leilao {i}')
139 coverage_positive, \
140 average_width, \
141 rmse, \
142 positive_rmse, \
143 mean_lance, \
144     = performance(y_std*y_test, y_std*yest50, y_std*yest5,
145                   y_std*yest95 )
146 coverage_positive_i.append( coverage_positive )
147 average_width_i.append( average_width )
148 rmse_i.append( rmse )
149 positive_rmse_i.append( positive_rmse )
150 mean_lance_i.append( mean_lance )
```