

UNIVERSIDADE FEDERAL DE SANTA MARIA
CAMPUS CACHOEIRA DO SUL
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Jaderson Rosa dos Santos

**ANÁLISE ESTATÍSTICA PARA SELEÇÃO DOS ATRIBUTOS QUE
CARACTERIZAM OS INDICADORES DE CONTINUIDADE DO
FORNECIMENTO DE ENERGIA ELÉTRICA**

Cachoeira do Sul, RS
2021

Jaderson Rosa dos Santos

ANÁLISE ESTATÍSTICA PARA SELEÇÃO DOS ATRIBUTOS QUE CARACTERIZAM OS INDICADORES DE CONTINUIDADE DO FORNECIMENTO DE ENERGIA ELÉTRICA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia Elétrica da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Engenharia Elétrica**. Defesa realizada por videoconferência.

ORIENTADORA: Prof.^a Laura Lisiane Callai dos Santos

Cachoeira do Sul, RS
2021

©2021

Todos os direitos autorais reservados a Jaderson Rosa dos Santos. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

End. Eletr.: r.jaderson@gmail.com

Jaderson Rosa dos Santos

ANÁLISE ESTATÍSTICA PARA SELEÇÃO DOS ATRIBUTOS QUE CARACTERIZAM OS INDICADORES DE CONTINUIDADE DO FORNECIMENTO DE ENERGIA ELÉTRICA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia Elétrica da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Engenharia Elétrica**.

Aprovado em 25 de agosto de 2021:

Laura Lisiane Callai dos Santos, Dr. (UFSM)
(Presidente/Orientadora)

Maicon Jaderson Silveira Ramos, Dr. (UFRGS) (videoconferência)

Nelson Knak Neto, Dr. (UFSM) (videoconferência)

Cachoeira do Sul, RS
2021

NUP: 23081.069134/2021-21

Prioridade: Normal

Homologação de Ata

010 - Organização e Funcionamento

COMPONENTE

Ordem	Descrição	Nome do arquivo
1	Ata de defesa de trabalho de conclusão de curso (TCC) (125.322)	Aprovacao_Jaderson.pdf

Assinaturas

26/08/2021 11:25:55

LAURA LISIANE CALLAI DOS SANTOS (PROFESSOR DO MAGISTÉRIO SUPERIOR)

31.04.00.00.0.0 - COORDENAÇÃO ACADÊMICA - UFSM-CS - C_ACA_CampusCS

26/08/2021 11:41:25

NELSON KNAK NETO (PROFESSOR DO MAGISTÉRIO SUPERIOR)

31.04.00.00.0.0 - COORDENAÇÃO ACADÊMICA - UFSM-CS - C_ACA_CampusCS

27/08/2021 12:09:41

MAICON JADERSON SILVEIRA RAMOS (Pessoa Física)

Usuário Externo (001.***.***.**))



Código Verificador: 810122

Código CRC: a5ea7691

Consulte em: <https://portal.ufsm.br/documentos/publico/autenticacao/assinaturas.html>



AGRADECIMENTOS

Agradeço aos meus pais por todo apoio, incentivo e esforço incondicional ao longo destes anos.

À minha namorada Amanda, por todo companheirismo, carinho, incentivo e compreensão.

À professora Laura, pela confiança em mim depositada e pela dedicação durante este trabalho.

Aos meus colegas e amigos, pela parceria e troca de aprendizados durante a graduação.

Aos professores da UFSM campus Cachoeira do Sul, por todo conhecimento compartilhado.

RESUMO

ANÁLISE ESTATÍSTICA PARA SELEÇÃO DOS ATRIBUTOS QUE CARACTERIZAM OS INDICADORES DE CONTINUIDADE DO FORNECIMENTO DE ENERGIA ELÉTRICA

AUTOR: Jaderson Rosa dos Santos

ORIENTADORA: Laura Lisiane Callai dos Santos

A qualidade da energia elétrica (QEE) fornecida pelas distribuidoras de energia elétrica é regulada pela Agência Nacional de Energia Elétrica (ANEEL) em 3 aspectos principais: qualidade do serviço, qualidade do produto e qualidade do atendimento. No que tange a qualidade do serviço, são estabelecidos indicadores de continuidade do serviço e limites para estes indicadores. A metodologia vigente de definição dos limites está descrita na Nota Técnica nº102/2014-SRD/ANEEL. Esta metodologia é constantemente atualizada visto que o Brasil possui um extenso território e, por conseguinte, determinar as variáveis que relacionam-se com as interrupções do fornecimento de energia torna-se uma tarefa não trivial. Nesse contexto, o presente trabalho tem como objetivo analisar a mudança dos critérios da análise estatística para seleção dos atributos que caracterizam os indicadores de continuidade. A análise proposta baseia-se nos mesmos procedimentos utilizados pela ANEEL para a definição do modelo atual para caracterização dos indicadores de continuidade, no entanto, são utilizadas técnicas estatísticas distintas das empregadas pela ANEEL. Para validar a metodologia proposta, foram desenvolvidos oito estudos de casos utilizando o *software SAS University Edition*. Nestes estudos de caso, avaliou-se a mudança do limite de correlação e do fator de inflação de variância (VIF), assim como a alteração da técnica de regressão. Verificou-se que, utilizando a regressão *Stepwise*, o aumento no limite de correlação tende a diminuir o poder explicativo e a quantidade de variáveis nos modelos, no entanto, aumenta o poder explicativo parcial de cada variável presente nos modelos.

Palavras-chave: Qualidade. Indicadores. Atributos. Limites. Continuidade.

ABSTRACT

STATISTICAL ANALYSIS FOR THE SELECTION OF ATTRIBUTES THAT CHARACTERIZE THE INDICATORS OF CONTINUITY OF ELECTRIC ENERGY SUPPLY

AUTHOR: Jaderson Rosa dos Santos

ADVISOR: Laura Lisiane Callai dos Santos

The quality of electricity (QEE) supplied by electricity distributors is regulated by the National Electric Energy Agency (ANEEL) in 3 main aspects: service quality, product quality and service quality. With regard to service quality, service continuity indicators and limits for these indicators are established. The current methodology for defining the limits is described in Technical Note nº102/2014-SRD/ANEEL. This methodology is constantly updated since Brazil has an extensive territory and, therefore, determining the variables that are related to interruptions in energy supply becomes a non-trivial task. In this context, the present work aims to analyze the change in the statistical analysis criteria for selecting the attributes that characterize the continuity indicators. The proposed analysis is based on the same procedures used by ANEEL to define the current model for characterizing continuity indicators, however, statistical techniques different from those employed by ANEEL are used. To validate the proposed methodology, eight case studies were developed using the software SAS University Edition. In these case studies, the change in the correlation limit and the variance inflation factor (VIF) were evaluated, as well as the change in the regression technique. It was found that, using Stepwise regression, the increase in the correlation threshold tends to decrease the explanatory power and the number of variables in the models, however, it increases the partial explanatory power of each variable present in the models.

Keywords: Quality. Indicators. Attributes. Limits.

LISTA DE FIGURAS

Figura 2.1 – Determinação do limite do indicador DEC para um conjunto que pertence ao SIN.	19
Figura 2.2 – Definição dos Modelos de DEC e FEC.	21
Figura 2.3 – Comparação entre os modelos de DEC e FEC.	22
Figura 2.4 – Modelo de Regressão Hipotético.	27
Figura 2.5 – Modelo Sobreajustado.	29
Figura 2.6 – Fluxograma de Determinação de λ	31
Figura 2.7 – Fluxograma dos Procedimentos do Método <i>Stepwise</i>	35
Figura 2.8 – Representação Gráfica do Cálculo do R^2	36
Figura 3.1 – Metodologia Proposta.	40
Figura 3.2 – Definição do modelo.	42
Figura 4.1 – Modelos do Caso 1	48
Figura 4.2 – Quantidade de Atributos por Modelo	60
Figura 4.3 – Poder Explicativo por Modelo	61
Figura 4.4 – Variação do Poder Explicativo por Modelo em Relação ao Caso 1	62
Figura 4.5 – Poder Explicativo por Variável	62
Figura 4.6 – Variação do Poder Explicativo por Variável em Relação ao Caso 1	63

LISTA DE TABELAS

Tabela 4.1 – Parâmetros utilizados no estudo de caso 1.....	44
Tabela 4.2 – Resultado da regressão <i>Stepwise</i> para o modelo do DEC considerando e desconsiderando variáveis socioeconômicas.....	45
Tabela 4.3 – Modelos Selecionados para o DEC.....	46
Tabela 4.4 – Resultado da regressão <i>Stepwise</i> para o modelo do FEC considerando e desconsiderando variáveis socioeconômicas.....	47
Tabela 4.5 – Modelos Selecionados para o FEC.....	47
Tabela 4.6 – Parâmetros utilizados no estudo de caso 2.....	49
Tabela 4.7 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 2.....	50
Tabela 4.8 – Parâmetros utilizados no estudo de caso 3.....	51
Tabela 4.9 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 3.....	51
Tabela 4.10 – Parâmetros utilizados no estudo de caso 4.....	52
Tabela 4.11 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 4.....	52
Tabela 4.12 – Parâmetros utilizados no estudo de caso 5.....	53
Tabela 4.13 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 5.....	54
Tabela 4.14 – Parâmetros utilizados no estudo de caso 6.....	55
Tabela 4.15 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 6.....	55
Tabela 4.16 – Parâmetros utilizados no estudo de caso 7.....	56
Tabela 4.17 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 7.....	57
Tabela 4.18 – Parâmetros utilizados no estudo de caso 8.....	58
Tabela 4.19 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 8.....	59
Tabela 4.20 – Parâmetros utilizados nos estudos de caso.....	60
Tabela A.1 – 146 Atributos.....	69

LISTA DE QUADROS

Quadro 2.1 – Variáveis Seleccionadas para o Modelo da ANEEL.	23
---	----

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
DEC	Duração Equivalente de Interrupção por Unidade Consumidora
DIC	Duração de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão
DICRI	Duração da Interrupção Individual Ocorrida em Dia Crítico por Unidade Consumidora ou por Ponto de Conexão
DMIC	Duração Máxima de Interrupção Contínua por Unidade Consumidora ou por Ponto de Conexão
DNAEE	Departamento Nacional de Águas e Energia
FEC	Frequência de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão
FIC	Frequência de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão
MMQO	Método dos Mínimos Quadrados Ordinários
PRODIST	Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico
SIN	Sistema Interligado Nacional
SIGR	Sistema de Informação Geográfica Regulatório
SQR	Soma dos Quadrados dos Resíduos
STQ	Soma Total de Quadrados
VIF	Fator de Inflação de Variância

LISTA DE SÍMBOLOS

t	Tempo de interrupção
t_{max}	Tempo de interrupção máximo
Cc	Número total de unidades consumidoras faturadas por conjunto
ρ	Coefficiente de correlação de Pearson
x_i	Amostra da variável x
\bar{x}	Média Aritmética da variável x
y_i	Amostra da variável y
\bar{y}	Média Aritmética da variável y
R	Coefficiente de correlação de Spearman
d	Diferença entre o <i>ranking</i> da variável x_i e y_i
y	Variável independente
x	Variável dependente
β	Parâmetro
ε	Resíduo

SUMÁRIO

1	INTRODUÇÃO	13
1.1	CONTEXTUALIZAÇÃO	13
1.2	MOTIVAÇÕES E JUSTIFICATIVA	14
1.3	OBJETIVOS	14
1.3.1	Objetivo Geral	14
1.3.2	Objetivos Específicos	15
1.4	ORGANIZAÇÃO DO TRABALHO	15
2	REVISÃO BIBLIOGRÁFICA	16
2.1	INDICADORES DE CONTINUIDADE DO SERVIÇO	16
2.1.1	Indicadores Individuais	16
2.1.2	Indicadores Coletivos	18
2.1.3	Metodologia Vigente para Definição dos Limites de Continuidade	18
2.1.4	Definição do Modelo Atual para Caracterização dos Indicadores DEC e FEC	20
2.2	COEFICIENTES DE CORRELAÇÃO	24
2.3	ANÁLISE DE REGRESSÃO	25
2.3.1	Método dos Mínimos Quadrados Ordinários	28
2.3.2	Regressão Ridge	30
2.3.3	Regressão LASSO	31
2.3.4	Métodos Stepwise	33
2.3.5	Coeficiente de Determinação e Coeficiente de Determinação Ajustado	36
2.3.6	Fator de Inflação de Variância	37
2.3.7	Considerações Finais	38
3	METODOLOGIA	39
3.1	SAS UNIVERSITY EDITION	42
3.2	CONSIDERAÇÕES FINAIS	43
4	ESTUDO DE CASO	44
4.1	ESTUDO DE CASO 1	44
4.1.1	DEC	44
4.1.2	FEC	46
4.1.3	Resultados do Estudo de Caso 1	48
4.2	ESTUDO DE CASO 2	49
4.3	ESTUDO DE CASO 3	50
4.4	ESTUDO DE CASO 4	52
4.5	ESTUDO DE CASO 5	53
4.6	ESTUDO DE CASO 6	54
4.7	ESTUDO DE CASO 7	56
4.8	ESTUDO DE CASO 8	58
4.9	COMPARAÇÃO DOS MODELOS	60
4.10	CONSIDERAÇÕES FINAIS	63
5	CONCLUSÃO	65
5.1	TRABALHOS FUTUROS	65
5.2	TRABALHOS PUBLICADOS	66
5.2.1	Artigos publicados em anais de eventos	66
	REFERÊNCIAS BIBLIOGRÁFICAS	67
	ANEXO A – 146 ATRIBUTOS	69

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

O setor de distribuição de energia elétrica brasileiro é definido como um monopólio natural, uma vez que as distribuidoras têm exclusividade de distribuição de energia para todos os consumidores cativos que estão na sua área de concessão. Como forma de assegurar a qualidade do serviço fornecido aos consumidores, bem como simular um mercado competitivo, a Agência Nacional de Energia Elétrica (ANEEL), na função de órgão regulador, avalia a qualidade da energia elétrica fornecida pelas distribuidoras em três aspectos: qualidade do produto, qualidade da serviço e qualidade do atendimento (TANURE, 2000). De acordo com Pessanha (2006), a qualidade do serviço é o ponto mais importante, uma vez que está relacionada com as interrupções do fornecimento de energia.

O processo de avaliação da qualidade do serviço no Brasil iniciou com a edição da Portaria nº46/1978 pelo Departamento Nacional de Águas e Energia Elétrica (DNAEE), na qual foram definidos os indicadores de continuidade Duração Equivalente de Interrupção por Unidade Consumidora (DEC) e Frequência de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão (FEC). Entretanto, não foram estabelecidas punições pelas transgressões dos limites de continuidade.

Nos anos 90 houve uma reestruturação do setor elétrico brasileiro, marcada pela desverticalização de empresas de energia, privatização de distribuidoras e extinção do DNAEE, cujas atribuições foram outorgadas à ANEEL (KAGAN; ROBBA; SCHMIDT, 2009). A partir destes fatos, viu-se a necessidade de reformar a regulamentação da qualidade do serviço. Nesse sentido, a ANEEL emitiu a Resolução nº 024/2000 (ANEEL, 2000), que anulou a Portaria nº46/1978 e adicionou as mudanças necessárias para adequação da regulamentação ao novo contexto. A principal atualização trata-se da metodologia de definição dos limites de continuidade baseada em uma análise comparativa de conjuntos de unidades consumidoras, cujo princípio é que conjuntos similares devem ter desempenho equivalente. Além disso, foi instaurada a imputação de penalidades pela transgressão dos limites de continuidade.

Desde a publicação da Resolução nº024/2000, a metodologia para definição dos limites de continuidade sofreu revisões periódicas, concedidas pela ANEEL por intermédio de audiências públicas. As atualizações ocorreram no sentido dos métodos de agrupamento de conjuntos e nos atributos que caracterizam os conjuntos de unidades consumidoras. A metodologia vigente, estabelecida pela Nota Técnica nº102/2014 (ANEEL, 2014), utiliza 6 atributos elétricos, geográficos e pluviométricos para caracterizar os conjuntos e aplica o método dinâmico para o agrupamento de conjuntos semelhantes.

Frente ao exposto, o presente trabalho propõe uma análise de atributos que influ-

enciam a continuidade do serviço. Para tanto, é definida uma metodologia que, a partir dos 146 atributos utilizados pela ANEEL para desenvolver o modelo vigente, são gerados modelos estatísticos para os indicadores DEC e FEC. São utilizados aparatos estatísticos para o estudo de quais atributos possuem maior relevância em relação ao indicadores de continuidade coletivos.

1.2 MOTIVAÇÕES E JUSTIFICATIVA

O Brasil possui um extenso território e, por conseguinte, o país apresenta diferentes características geográficas ao longo de sua área. Portanto, determinar as variáveis que relacionam-se com as interrupções do fornecimento de energia torna-se uma tarefa não trivial, fato que justifica as atualizações que ocorreram na metodologia de definição dos limites de continuidade, visando torná-la condizente com as características das distribuidoras de energia elétrica.

Dessa forma, é preciso investigar quais são as variáveis que produzem efeitos significativos nos indicadores de continuidade e a partir destas modelar o comportamento dos indicadores, no sentido de caracterizá-los de forma condizente com a realidade. Estas variáveis devem compor um modelo parcimonioso, visto que há uma grande dificuldade em desenvolver um modelo teórico que compreende o nível de heterogeneidade das áreas de concessão das distribuidoras e a complexidade da atividade de distribuição de energia elétrica.

Além disso, o emprego de um modelo com atributos que não apresentem causalidade com as interrupções, pode gerar limites intangíveis ou aliviar os limites de forma equivocada para as distribuidoras, uma vez que este problema cria agrupamentos de conjuntos heterogêneos. Em outros termos, quando ocorre a comparação de conjuntos a partir de características que não estão relacionadas com as interrupções, há a possibilidade de aplicação de multas indevidas.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Este trabalho tem como objetivo geral realizar uma análise estatística para seleção dos atributos que caracterizam os indicadores de continuidade coletivos.

1.3.2 Objetivos Específicos

- Caracterizar a metodologia vigente para definição dos limites de continuidade.
- Identificar técnicas estatísticas para elaboração de modelo de predição de variáveis.
- Desenvolver modelos para caracterização dos indicadores Duração Equivalente de Interrupção por Unidade Consumidora e Frequência Equivalente de Interrupção por Unidade Consumidora.
- Analisar diferentes modelos para os indicadores Duração Equivalente de Interrupção por Unidade Consumidora e Frequência Equivalente de Interrupção por Unidade Consumidora.
- Comparar os resultados obtidos com o modelo utilizado pela ANEEL.

1.4 ORGANIZAÇÃO DO TRABALHO

Este trabalho está dividido em 5 capítulos, incluindo esse introdutório.

No capítulo 2 é apresentada a revisão bibliográfica sobre indicadores de continuidade do serviço, abordando a forma como são apurados, a definição dos limites e a definição do modelo dos indicadores DEC e FEC. Também são caracterizadas as ferramentas estatísticas coeficiente de correlação e análise de regressão, objetivando descrever as técnicas utilizadas no estabelecimento do modelo atual, assim como apresentar alternativas a estas técnicas.

No capítulo 3 é apresentada a metodologia proposta para o desenvolvimento deste trabalho. A metodologia proposta está dividida em cinco partes principais. A primeira etapa constitui-se em definir os dados de entrada a serem analisados e, posteriormente, realizar a seleção destes dados, visando manter na análise apenas os atributos relevantes. Após, é realizada a definição dos métodos de regressão a serem empregados. Durante a quarta etapa, aplica-se os métodos de regressão às variáveis restantes da filtragem por coeficientes de correlação. Por fim, são analisados os modelos gerados.

O capítulo 4 apresenta os estudos de caso e resultados, que foram construídos variando os parâmetros de correlação, VIF e a técnica de regressão. Foram construídos 8 estudos de caso e, desta forma, foi possível analisar o impacto da mudança dos parâmetros no poder explicativo e na quantidade de variáveis presentes nos modelos.

Por fim, no capítulo 5 são apresentadas as considerações finais do trabalho.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo será apresentada uma breve revisão da literatura dos objetos necessários para o desenvolvimento do trabalho. Primeiramente, são definidos os indicadores de continuidade do serviço, assim como a metodologia atual para estabelecimento do limite destes. Após, é apresentado o conceito de coeficientes de correlação, assim como as técnicas coeficiente de correlação de Pearson e coeficiente de correlação de Spearman, utilizados pela ANEEL para analisar quais variáveis possuem alta ou baixa correlação com DEC e FEC. Por fim, é abordada a análise de regressão, onde é caracterizada a regressão *Stepwise* aplicada para a construção do modelo atual de DEC e FEC. Além disso, são apresentadas as técnicas de regressão LASSO e Ridge.

2.1 INDICADORES DE CONTINUIDADE DO SERVIÇO

Os indicadores de continuidade avaliam, essencialmente, dois aspectos da interrupção do fornecimento de energia elétrica: duração e frequência. De acordo com Tanure (2004), "A duração das interrupções está intimamente ligada aos meios humanos e materiais empregados para a recomposição e reparo da rede, bem como às facilidades existentes para se recuperar o sistema após cada interrupção [...]". Quanto à frequência de interrupção, Tanure (2004, p.19) afirma que "A frequência das interrupções caracteriza a fragilidade do sistema frente ao meio ambiente (causas externas) e a degradação do sistema por envelhecimento e/ou falta de manutenção adequada (causas internas) [...]."

Os cálculos para apuração dos indicadores coletivos e individuais, assim como os critérios para a formação de conjuntos de unidades consumidoras, são definidos no módulo 8 dos Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico (PRODIST) (ANEEL, 2021). Além disso, o módulo 8 do PRODIST estabelece que os indicadores devem ser apurados mensalmente, trimestralmente e anualmente.

2.1.1 Indicadores Individuais

Os indicadores individuais são calculados por unidade consumidora. Na subseção 5.4 da seção 8.2 do módulo 8 do PRODIST(ANEEL, 2021), são apresentadas as equações para a apuração dos seguintes indicadores:

- Duração de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão (DIC), que indica o tempo acumulado que, durante o período de apuração, cada unidade consumidora teve o fornecimento de energia elétrica interrompido. O cálculo

desse indicador é dado pela Equação 2.1.

$$DIC = \sum_{i=1}^n t(i) \quad (2.1)$$

- Frequência de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão (FIC), que indica o número de interrupções que ocorreram durante o período de apuração para cada unidade consumidora. O cálculo desse indicador é dado pela Equação 2.2.

$$FIC = n \quad (2.2)$$

- Duração Máxima de Interrupção Contínua por Unidade Consumidora ou por Ponto de Conexão (DMIC), que indica o tempo máximo que uma unidade consumidora ficou desabastecida de energia elétrica durante o período de apuração. O cálculo desse indicador é dado pela Equação 2.3.

$$DMIC = t(i)_{max} \quad (2.3)$$

- Duração da Interrupção Individual ocorrida em Dia Crítico por unidade consumidora ou por ponto de conexão (DICRI), que indica o tempo que, em Dia Crítico durante o período de apuração, cada unidade consumidora teve o fornecimento de energia elétrica interrompido. O cálculo desse indicador é dado pela Equação 2.4.

$$DICRI = t_{critico} \quad (2.4)$$

Em que:

i: índice de interrupções da unidade consumidora ou por ponto de conexão no período de apuração, variando de 1 a n;

n: número de interrupções da unidade consumidora ou por ponto de conexão considerado, no período de apuração;

t(i): tempo de duração da interrupção (i) da unidade consumidora considerada ou do ponto de conexão, no período de apuração;

t(i)max: valor correspondente ao tempo da máxima duração de interrupção contínua (i), no período de apuração, verificada na unidade consumidora ou no ponto de conexão considerado, expresso em horas e centésimos de horas.

Os indicadores individuais são a base do cálculo de apuração dos indicadores coletivos.

2.1.2 Indicadores Coletivos

Os indicadores coletivos são calculados para cada conjunto de unidades consumidoras. Na subseção 5.5 da seção 8.2 do módulo 8 do PRODIST (ANEEL, 2021), são apresentadas as equações para a apuração dos seguintes indicadores coletivos:

- Duração Equivalente de Interrupção por Unidade Consumidora (DEC), que indica o tempo médio que, durante o período de apuração, cada unidade consumidora teve o fornecimento de energia elétrica interrompido. O cálculo desse indicador é dado pela Equação 2.5.

$$DEC = \frac{\sum_{i=1}^{C_c} DIC(i)}{C_c} \quad (2.5)$$

- Frequência de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão (FEC), que indica a quantidade média de vezes que ocorreram interrupções durante o período de apuração para cada unidade consumidora. O cálculo desse indicador é dado pela Equação 2.6.

$$FEC = \frac{\sum_{i=1}^{C_c} FIC(i)}{C_c} \quad (2.6)$$

Em que:

i: índice de unidades consumidoras atendidas em BT ou MT faturadas do conjunto;

C_c: número total de unidades consumidoras faturadas do conjunto no período de apuração, atendidas em BT ou MT;

DIC(i): Duração de Interrupção Individual por Unidade Consumidora, excluindo-se as centrais geradoras;

FIC(i): Frequência de Interrupção Individual por Unidade Consumidora, excluindo-se as centrais geradoras.

A apuração dos indicadores coletivos é necessária para avaliar o desempenho dos conjuntos em relação aos limites, que são definidos por meio de uma análise comparativa de desempenho.

2.1.3 Metodologia Vigente para Definição dos Limites de Continuidade

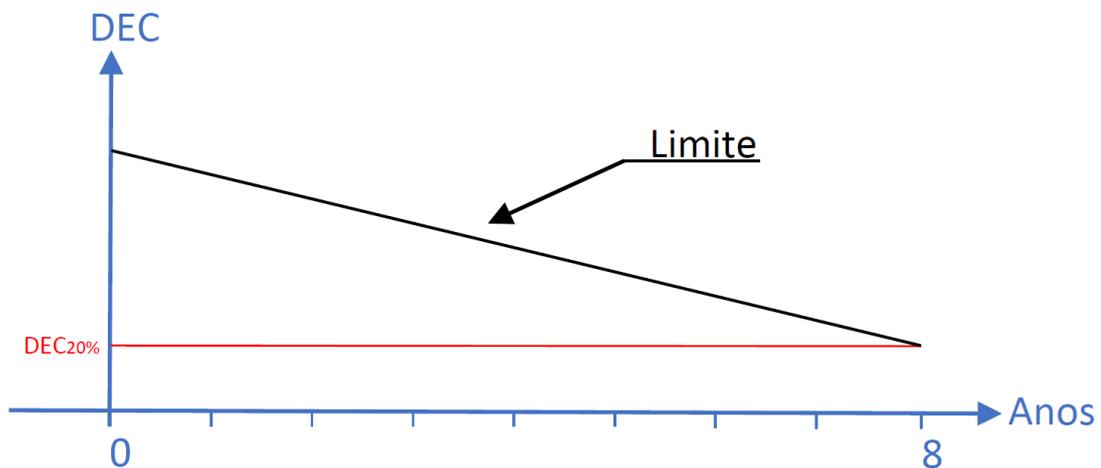
A ANEEL regula e controla o desempenho das distribuidoras através de limites estabelecidos para os indicadores de continuidade. As distribuidoras devem realizar compensações por transgressões nos limites individuais, ou seja, a restituição ocorre para cada unidade consumidora cujo serviço contratado não atingiu as metas de continuidade.

A metodologia vigente para definição dos limites está descrita na Nota Técnica nº102/2014-SRD/ANEEL (ANEEL, 2014), que apresenta as mudanças aplicadas à metodologia anterior, retratada na Nota Técnica nº 021/2011-SRD/ANEEL, assim como a ponderação das contribuições enviadas à Audiência Pública nº 029/2014, cujo tema foi a revisão da metodologia de definição dos indicadores DEC e FEC.

A metodologia de análise comparativa utilizada pela ANEEL baseia-se na técnica de regulação por desempenho *Yardstick Competition*, proposta por Tanure (TANURE, 2004). Esse método de definição dos limites parte da premissa de que conjuntos de unidades consumidoras semelhantes devem possuir desempenhos semelhantes. Neste modelo de regulação, os conjuntos devem alcançar a referência do agrupamento de conjuntos, também chamado de *cluster*.

Na metodologia vigente, utiliza-se o percentil 20 como referência para conjuntos pertencentes ao sistema interligado nacional (SIN) e o percentil 50 para conjuntos isolados. O limite é dado por uma equação linear, cujo ponto inicial é o limite do *cluster* no ano em que é calculado (ano 0) e o ponto final é o desempenho da referência do *cluster*, que deve ser alcançado 8 anos (ano 8) após o ano 0 (ANEEL, 2014). A Figura 2.1 ilustra a determinação do limite do indicador DEC para um conjunto que pertence ao SIN, ou seja, a referência do *cluster* é o segundo decil, cujo limite apurado no ano 0 deve ser atingido pelos demais conjuntos no ano 8.

Figura 2.1 – Determinação do limite do indicador DEC para um conjunto que pertence ao SIN.



Fonte: Adaptado de ANEEL (2010).

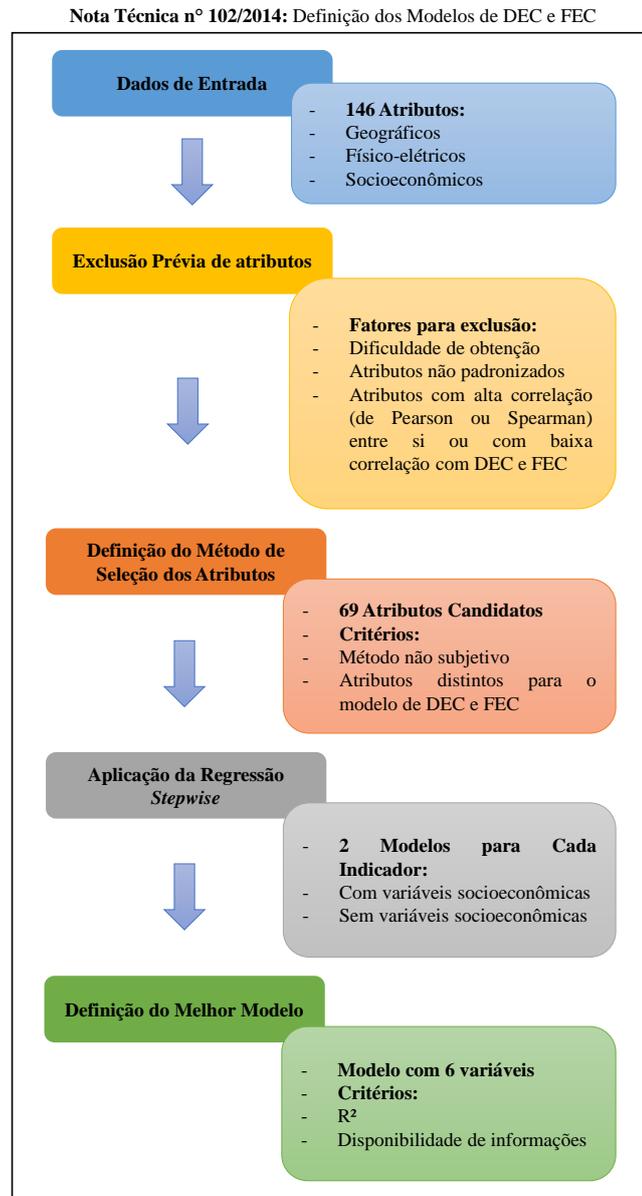
Para a formação dos *clusters*, elaborou-se um modelo para a caracterização dos conjuntos com base nos indicadores DEC e FEC. Por meio deste modelo, pode-se realizar o agrupamento dos conjuntos semelhantes, utilizando a técnicas de clusterização dinâmica, onde para cada conjunto é formado um agrupamento. Para cada agrupamento é definido um conjunto de referência e a partir da referência é estabelecido o limite do

cluster.

2.1.4 Definição do Modelo Atual para Caracterização dos Indicadores DEC e FEC

As etapas para definição do modelo atual, realizadas pela ANEEL, estão ilustradas na Figura 2.2. Como variáveis explicativas, a ANEEL definiu 146 atributos geográficos, físico-elétricos, socioeconômicos, meteorológicos e de vegetação provenientes do Sistema de Informação Geográfica Regulatório (SIG-R). Procedeu-se uma exclusão prévia de atributos, objetivando-se eliminar possíveis fontes de erro à análise. Os critérios utilizados para a retirada de atributos foram: dificuldade de obtenção, atributos não padronizados, atributos com alta correlação entre si ou com baixa correlação com DEC e FEC, atributos não atualizados e atributos redundantes. A avaliação da correlação entre os atributos deu-se pelo coeficiente de Pearson, onde, dos pares de atributos que apresentaram correlação maior que 0,9, manteve-se o que possui maior correlação com DEC e FEC. Além disso, utilizou-se a correlação mínima de 0,2 (Pearson ou Spearman) com DEC e FEC para manter os atributos para a próxima etapa (ANEEL, 2014).

Figura 2.2 – Definição dos Modelos de DEC e FEC.



Fonte: Adaptado de ANEEL (2014)

Após a eliminação prévia, norteadas pelos critérios citados anteriormente, restaram 69 atributos candidatos. A próxima etapa consistiu em definir o método de seleção dos atributos. Balizada por contribuições apresentadas à Audiência Pública n° 029/2014, nas quais sugeriu-se a substituição da análise fatorial, então utilizada, devido à subjetividade inerente a esta técnica, definiu-se a como método de seleção dos atributos a Regressão *Stepwise* (ANEEL, 2014).

Antes de proceder ao método de seleção de atributos, os dados foram normalizados

objetivando uma distribuição mais próxima da normal, de acordo com a Equação 2.7.

$$X_T = \ln(1 + X) \quad (2.7)$$

Em que:

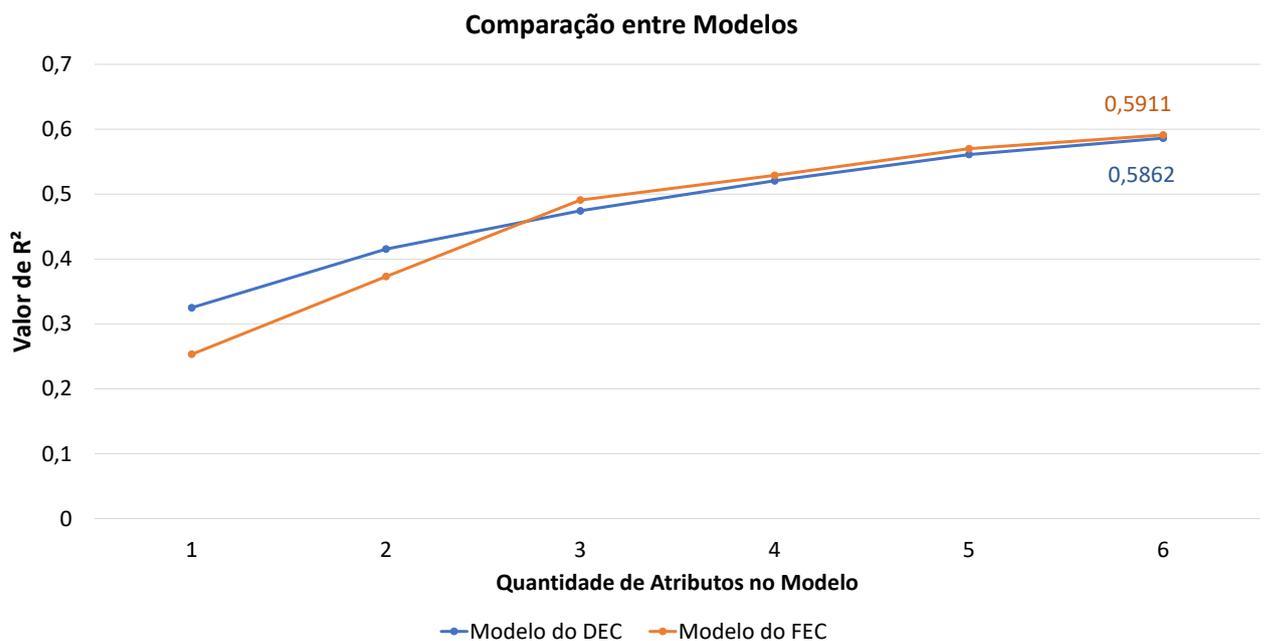
X_T : variável transformada;

X: variável original.

Tendo em posse os atributos normalizados e o aparato estatístico para escolha das variáveis do modelo, procedeu-se à eliminação de atributos por multicolinearidade. Em outras palavras, foi eliminada a redundância entre variáveis. A métrica utilizada foi o fator de inflação de variância, ou *Variance Inflation Factor* (VIF), optando-se por um VIF máximo igual a 5, considerando que este valor já elimina a multicolinearidade, para que a variável passe para a próxima etapa. Além disso, foram elaborados dois modelos para cada indicador, considerando as variáveis socioeconômicas e desconsiderando-as (ANEEL, 2014).

Os atributos restantes da eliminação por multicolinearidade foram submetidos à Regressão *Stepwise*, visando determinar o modelo final com maior poder explicativo para os indicadores. Para tanto, utilizou-se o coeficiente de determinação R^2 para avaliar e definir o modelo apropriado. Os resultados obtidos são mostrados na Figura 2.3, onde é possível visualizar que os modelos do DEC e do FEC possuem R^2 igual a 0,58 e 0,59, respectivamente, isto denota que estes modelos explicam, aproximadamente, 59% das variações de DEC e FEC.

Figura 2.3 – Comparação entre os modelos de DEC e FEC.



A ANEEL definiu o modelo com 6 variáveis desconsiderando as socioeconômicas para descrever os indicadores, utilizando como justificativa o fato de que, a partir de 6 atributos, há pouca variação do poder explicativo em função do acréscimo de novas variáveis. Além disso, optou-se pelo modelo sem as variáveis socioeconômicas pois a atualização desses dados ocorre decenalmente, isto é, pode haver uma defasagem temporal entre as informações utilizadas. Os atributos selecionados para os modelos de DEC e FEC são mostrados no Quadro 2.1. Cabe ressaltar que os modelos diferem apenas por 1 atributo, ou seja, 5 atributos são comuns a ambos os modelos.

Quadro 2.1 – Variáveis Selecionadas para o Modelo da ANEEL.

DEC		FEC	
Sigla	Atributo	Sigla	Atributo
PC_NUC_URB	Percentual de Número de Unidades consumidoras por localização Urbana (%)	PC_VRAM	Percentual de Área com Vegetação Remanescente Alta ou Média (%)
PC_VRAM	Percentual de Área com Vegetação Remanescente Alta ou Média (%)	PLUV	Precipitação Pluviométrica Média Anual (mm)
PC_ERMT_3F	Percentual de Redes Média Tensão Trifásicas (%)	PC_NUC_URB	Percentual de Número de Unidades consumidoras por localização Urbana (%)
PLUV	Precipitação Pluviométrica Média Anual (mm)	NUC_COM	Número de Unidades Consumidoras da Classe Comercial
CM_NUC_RES	Consumo Médio por Unidade Consumidora da Classe Residencial	CM_NUC_RES	Consumo Médio por Unidade Consumidora da Classe Residencial
NUC_IND	Número de Unidades Consumidoras da Classe Industrial	PC_ERMT_3F	Percentual de Redes Média Tensão Trifásicas (%)

Fonte: Adaptado de ANEEL (2014).

Na próxima seção, são retratados métodos estatísticos para elaboração de mo-

delos de predição de variáveis, visando caracterizar as técnicas empregadas pela ANEEL, bem como identificar problemas inerentes a estes métodos. Além disso, são apresentadas outras técnicas de análise de regressão com características distintas da regressão empregada pela ANEEL.

2.2 COEFICIENTES DE CORRELAÇÃO

Os Coeficientes de Correlação são métodos estatísticos que quantificam a associação entre duas variáveis, isto é, definem o quanto uma variável influencia a outra. Esses coeficientes exprimem a relação entre as variáveis em uma faixa que varia de -1 a +1. Coeficientes de correlação iguais a 1 ou -1 são chamados de correlações perfeitas, todavia, em dados reais não existe correlação perfeita, apenas correlação forte, isto é, próxima a 1 ou -1 (SCHIELD, 1995).

Quando o coeficiente de correlação é negativo, indica que as variáveis são inversamente correlacionadas, ou seja, na medida em que uma variável aumenta, a outra diminui. Já um coeficiente de correlação positivo demonstra que as variáveis analisadas variam proporcionalmente (SCHIELD, 1995).

Uma correlação forte entre variáveis não implica que uma variável causou a outra, isto significa que a análise de correlação não descreve a causalidade, apenas fornece indícios para tanto (SCHIELD, 1995). Além disso, esta técnica não diferencia as variáveis em dependente ou independente, a análise ocorre de forma simétrica (GUJARATI; PORTER, 2011).

Existem diversos coeficientes de correlação lineares ou não lineares, dentre eles estão o Coeficiente de Correlação de Pearson e o Coeficiente de Correlação de Spearman, que foram utilizados pela ANEEL para realizar uma exclusão prévia de atributos durante o processo de definição dos modelos de DEC e FEC (ANEEL, 2014).

O coeficiente de correlação de Pearson determina a relação linear entre duas variáveis e é independente de unidades de medida. Esta técnica é significativamente afetada por valores extremos, que podem distorcer o valor da correlação. Portanto, é necessário que as amostras analisadas estejam normalmente distribuídas (HEUMANN; SCHOMAKER et al., 2016). O coeficiente de correlação de Pearson entre variáveis x e y é dado pela Equação 2.8.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.8)$$

Em que:

- ρ : coeficiente de correlação de Pearson;
- x_i : amostra da variável x ;

- \bar{x} : média aritmética da variável x ;
- y_i : amostra da variável y ;
- \bar{y} : média aritmética da variável y ;
- n : quantidade de amostras.

O coeficiente de correlação de Spearman define a relação não linear entre duas variáveis. Este método não utiliza diretamente a magnitude das amostras das variáveis para determinar o grau de associação entre elas, mas sim o *ranking* delas (HEUMANN; SCHOMAKER et al., 2016). Isto é, as amostras precisam apenas estarem ordenadas de acordo com suas magnitudes. A Equação 2.9 é utilizada para obtenção do coeficiente de correlação de Spearman.

$$R = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2.9)$$

Em que:

- R : coeficiente de correlação de Spearman;
- d : diferença entre o *ranking* da variável x_i e y_i ;
- n : quantidade de amostras.

Uma vez que o coeficiente de correlação de Spearman não utiliza a magnitude das variáveis, pode ser empregado tanto para variáveis contínuas quanto para variáveis ordinais. Além disso, este coeficiente é menos influenciado por valores extremos (HEUMANN; SCHOMAKER et al., 2016).

Apesar de os coeficientes de correlação não modelarem a relação entre as variáveis, eles podem ser utilizados como critérios de exclusão de variáveis em casos onde se deseja eliminar variáveis altamente correlacionadas entre si. Para a modelagem da associação entre variáveis, utilizam-se técnicas de regressão.

2.3 ANÁLISE DE REGRESSÃO

A análise de regressão tem sido amplamente empregada em diversas áreas científicas. Diferentemente dos coeficientes de correlação, que determinam apenas a intensidade da relação entre duas variáveis, a regressão permite analisar e modelar a relação entre uma variável dependente, também denominada de variável resposta, e uma ou mais variáveis independentes, também chamadas de variáveis preditoras, que fornecem informações a respeito da variável resposta (MONTGOMERY; PECK; VINING, 2012, p. 1). O

objetivo da regressão é estimar um modelo capaz de prever o valor médio de uma variável com base em valores conhecidos (amostras) de outras variáveis (GUJARATI; PORTER, 2011, p. 39).

Todos os modelos de regressão envolvem constantes desconhecidas, chamadas de parâmetros, que regem o comportamento das variáveis independentes em relação ao modelo. Os parâmetros são determinados a partir dos dados utilizados na regressão (RAWLINGS; PANTULA; DICKEY, 2011, p. 2). Em outros termos, a análise de regressão resulta em uma equação que permite fazer previsões da variável resposta a partir de alterações na variável preditora. Desta forma, pode-se representar o modelo de regressão linear simples pela Equação 2.10.

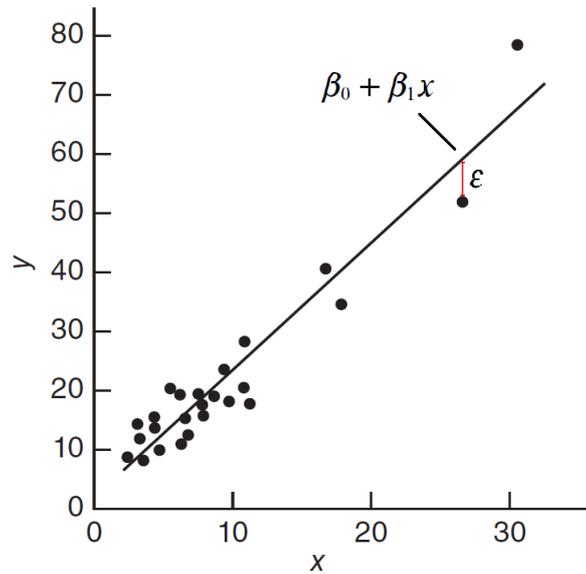
$$y = \beta_0 + \beta_1 x_1 + \varepsilon \quad (2.10)$$

Em que:

- y : variável independente;
- x_1 : variável dependente;
- β_0 e β_1 : parâmetros;
- ε : resíduo.

O resíduo ε da Equação 2.10 é uma variável aleatória que exprime a diferença entre os valores da reta gerada pela regressão e os valores observados da variável resposta (MONTGOMERY; PECK; VINING, 2012). Em outras palavras, a variável ε expressa o desvio de uma amostra de y em relação à média de todas as amostras de y (RAWLINGS; PANTULA; DICKEY, 2011). A Figura 2.4 mostra uma representação gráfica de um modelo de regressão hipotético, onde pode-se observar a reta do modelo ($\beta_0 + \beta_1 x$) e o resíduo, dado pela diferença entre as observações da resposta e a reta do modelo.

Figura 2.4 – Modelo de Regressão Hipotético.



Fonte: Adaptado de Montgomery, Peck e Vining (2012).

De acordo com Gujarati e Porter (2011), uma equação pode ser classificada como linear no que tange as variáveis ou em relação aos parâmetros. Uma função é dita linear nas variáveis se a variável dependente estiver elevada apenas à primeira potência e não estiver sendo multiplicada ou dividida por outra variável. O mesmo ocorre para uma função linear nos parâmetros, onde os parâmetros devem ter expoente 1 e não podem ser multiplicados ou divididos por outros parâmetros (GUJARATI; PORTER, 2011, p. 62). Em vista disso, a Equação 2.10 pode ser classificada como linear nas variáveis e nos parâmetros. No entanto, em análise de regressão utiliza-se o termo linear apenas para equações lineares nos parâmetros.

Além disso, Equação 2.10 representa um modelo de regressão linear simples pois há apenas uma variável preditora, entretanto, em casos reais a variável resposta, no geral, é explicada por diversas variáveis regressoras, o que caracteriza uma regressão linear múltipla, dada pela Equação 2.11.

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i + \epsilon \quad (2.11)$$

Em que:

- i : índice da variável independente, que varia de 1 até n , em que n é a quantidade de variáveis independentes.

Em análise de regressão linear múltipla, além de estimar os parâmetros das variáveis predictoras, objetiva-se determinar a quantidade ideal de variáveis para comporem o modelo, uma vez que um modelo com diversas variáveis torna-se difícil de interpretar. Para

tanto, existem diversas técnicas, por exemplo: Método dos Mínimos Quadrados Ordinários, regressão *ridge*, regressão LASSO e regressão *Stepwise*.

2.3.1 Método dos Mínimos Quadrados Ordinários

O Método dos Mínimos Quadrados Ordinários (MMQO) é uma técnica vastamente empregada para estimar os parâmetros das variáveis independentes devido a sua simplicidade matemática. Esta ferramenta utiliza a otimização matemática para minimizar a soma dos resíduos da equação de regressão, de tal forma que o modelo matemático da regressão seja o mais próximo possível dos valores observados (GUJARATI; PORTER, 2011). Para tanto, é utilizado o modelo de regressão em termos de amostras, expresso pela Equação 2.12 (MONTGOMERY; PECK; VINING, 2012).

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{ik} x_{ik} + \varepsilon_i \quad (2.12)$$

Em que:

- i : índice da amostra, que varia de 1 até n amostras;
- k : índice da variável independente, que varia de 1 até n variáveis independentes.

A função do MMQO é obtida isolando-se o resíduo da Equação 2.12 e elevando todos os termos ao quadrado. Além disso, utiliza-se o somatório de todas as observações, uma vez que o objetivo do MMQO é reduzir o resíduo em todo o conjunto de amostras (GUJARATI; PORTER, 2011). Desta forma, a equação do MMQO, que se deseja minimizar, é expressada de acordo com a Equação 2.13.

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (2.13)$$

Como consequência de os resíduos estarem elevados ao quadrado, resíduos maiores possuem pesos maiores, isto é, valores estimados da resposta distantes dos valores observados exercem maior influência sobre o MMQO. Além disso, os resíduos podem assumir valores positivos ou negativos, desta forma, a soma destes é um valor próximo à zero. Portanto, utilizar a soma dos quadrados dos resíduos (SQR) é uma solução para o problema anteriormente citado (GUJARATI; PORTER, 2011).

Para estimação dos parâmetros que minimizam a soma dos resíduos da equação de regressão, deriva-se a equação X em relação a cada coeficiente da variável dependente e iguala-se a zero, o que resulta em uma equação para cada parâmetro presente no modelo. Nota-se que quanto mais variáveis regressoras estiverem sendo analisadas, maior será o esforço para determinação dos coeficientes. Como alternativa para este problema, utiliza-

se a forma matricial da Equação 2.13 para simplificar a solução (RAWLINGS; PANTULA; DICKEY, 2001). Assim sendo, a representação matricial da Equação 2.13 é expressa pela Equação 2.14.

$$\varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta) \quad (2.14)$$

Em que:

- Y: vetor $n \times 1$ das observações da variável resposta;
- X: matriz $n \times 1$ dos parâmetros, ou coeficientes de regressão;
- ε : vetor $n \times 1$ dos resíduos.

Derivando-se a Equação 2.14 em relação aos parâmetros β , obtém-se a Equação 2.15 para definição coeficientes de regressão.

$$\beta = (X'X)^{-1}X'Y \quad (2.15)$$

Os parâmetros estimados pela Equação 2.15, tipicamente, são diferentes de zero. Portanto, a interpretação do modelo final pode ser difícil se houveram diversas variáveis regressoras e, além do mais, tende a ocorrer o sobreajuste, ou *overfitting* (HASTIE; TIBSHIRANI; WAINWRIGHT, 2015), e apresentar alta variância. Em outros termos, o modelo final do MMQO, quando há uma quantidade relativamente grande de variáveis independentes, ajusta-se com precisão para as amostras analisadas, no entanto, tem pouca capacidade de prever resultados futuros. Um exemplo hipotético de regressão sobreajustada é mostrado na Figura 2.5, onde observa-se que o modelo ajusta-se para todos os dados, todavia, há alta variância.

Figura 2.5 – Modelo Sobreajustado.



Para solucionar os problemas inerentes ao MMQO, pode-se utilizar previamente métodos de seleção de variáveis ou, então, métodos que realizem tanto a seleção de variáveis quanto a definição dos parâmetros como, por exemplo, a regressão *ridge* e a regressão LASSO.

2.3.2 Regressão *Ridge*

A Regressão *Ridge* foi introduzida por Hoerl e Kennard (1970) como um aperfeiçoamento ao MMQO. Neste método, utiliza-se uma normalização L2 para limitar a minimização da SQR, objetivando obter um modelo mais parcimonioso, isto é, a regressão não se ajusta a todos os dados observados, entretanto, tem maior capacidade de prever resultados futuros. A equação da regressão *ridge*, que deseja-se minimizar, é representada pela Equação 2.16.

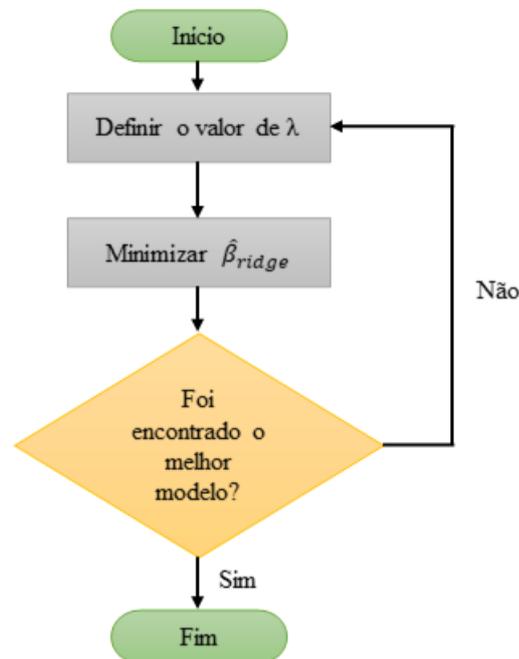
$$\hat{\beta}_{ridge} = \left[\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^k \beta_j^2 \right] \quad (2.16)$$

Em que:

- $\hat{\beta}_{ridge}$: coeficiente de regressão *ridge*;
- λ : parâmetro de ajuste do encolhimento.

A limitação da SQR, também chamada de penalidade, é controlada pela variável λ . Ao minimizar a Equação 2.16, nota-se que quanto maior o valor de λ , menor é o valor dos parâmetros, devido este coeficiente de controle estar multiplicando o quadrado dos parâmetros. Desta forma, quando λ for igual a zero, resta apenas a equação do MMQO; e quando tender a infinito, apenas o parâmetro β_0 possui relevância no modelo (HOERL; KENNARD, 1970), enquanto que os demais serão próximos de zero. Portanto, λ deve ser um valor maior que zero que reduza apenas as variáveis irrelevantes ao modelo. A Figura 2.6 mostra um fluxograma do método para obtenção do valor ideal de λ .

Figura 2.6 – Fluxograma de Determinação de λ .



Fonte: Adaptado de Hoerl e Kennard (1970)

Apesar de a regressão *ridge* não ser um método de seleção de variáveis por não anular variáveis, no trabalho de Luo e Liu (2017) esta técnica é utilizada para selecionar as variáveis previamente na confecção de um modelo de predição de áreas de fogo. Para tanto, Luo e Liu (2017) eliminam os parâmetros que foram reduzidos a valores próximos de zero e, posteriormente, utiliza a máquina de vetores de suporte para definir o modelo. Os resultados experimentais demonstram que o modelo final, utilizando a regressão *ridge* como seleção de variáveis, consegue prever áreas de fogo com eficiência.

No trabalho de Lin e Lin (2019), o MMQO e a regressão *ridge* são empregados para definir modelos que estimem a clorofila presente em folhas de árvores utilizando dados de refletância das cores amarelo e vermelho. Os resultados experimentais demonstraram que a regressão *ridge* obteve maior precisão de estimação que o MMQO. Além disso, Lin e Lin (2019) afirma que a regressão *ridge* obteve melhor performance por reduzir o erro de estimação.

Havendo necessidade de um método que selecione variáveis ao mesmo tempo que estipula os parâmetros, pode-se utilizar a regressão LASSO.

2.3.3 Regressão LASSO

A Regressão *Least Absolute Shrinkage and Selection Operator* (LASSO) foi apresentada por Tibshirani (1996) como forma de solucionar as deficiências apresentadas pe-

los métodos de seleção de modelos e pela regressão *ridge*. Os métodos de seleção de modelo eliminam variáveis por meio de processos automáticos que identificam redundâncias entre as variáveis regressoras, conseqüentemente, o modelo final é de fácil interpretação pois é reduzido. Entretanto, não é preciso, pois está sujeito à erros amostrais (TIBSHIRANI, 1996). Em relação à regressão Ridge, Tibshirani (1996) apontou como um método estável por reduzir os parâmetros de variáveis irrelevantes, no entanto, não define os coeficientes de regressão igual a zero, o que resulta em um modelo com diversas variáveis e, por conseguinte, de difícil interpretação.

A regressão LASSO, assim como a regressão ridge, é uma otimização do MMQO, entretanto, uma regularização L1 é aplicada à Equação 2.13 para limitar a soma do valor absoluto dos parâmetros. Desta forma, os parâmetros podem ser reduzidos a zero ou próximo de zero, fornecendo uma maneira automática de seleção de variáveis (HASTIE; TIBSHIRANI; WAINWRIGHT, 2015). A equação do método LASSO é expressa pela Equação 2.17.

$$\hat{\beta}_{lasso} = \left[\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^k |\beta_j| \right] \quad (2.17)$$

Em que:

- $\hat{\beta}_{lasso}$: coeficiente de regressão LASSO;
- λ : parâmetro de ajuste do encolhimento.

O problema de multicolinearidade também é solucionado aplicando-se a regressão LASSO. Se houverem variáveis preditoras com alta correlação entre si, este método mantém apenas uma e elimina as outras. Por consequência, obtém-se um modelo interpretável, isto é, com poucas variáveis e com alta precisão de previsão (MUTHUKRISHNAN; ROHINI, 2016).

Desde a introdução da regressão LASSO por Tibshirani (1996), diversos trabalhos vêm investigando este método. No trabalho de Muthukrishnan e Rohini (2016), são analisadas as técnicas de regressão MMQO, regressão *ridge* e regressão LASSO no que tange o ajuste do modelo e precisão de previsão, utilizando dados reais de pacientes com diabetes para modelar a relação da variável diabetes com outras variáveis regressoras. Os resultados experimentais demonstraram que a regressão LASSO apresentou a melhor seleção de variáveis e maior precisão de previsão. Posto isto, Muthukrishnan e Rohini (2016) concluíram que o método LASSO tem a melhor performance por reduzir parâmetros não significativos a zero.

Fan et al. (2015) apresentam uma comparação entre as técnicas a regressão LASSO, regressão LASSO adaptativo e Elastic Net em relação a predição e ajuste aos dados. Similarmente a Muthukrishnan e Rohini (2016), neste trabalho são utilizadas variáveis rela-

cionadas à diabete para modelar o comportamento desta doença. De acordo com os resultados, o método LASSO apresentou o melhor ajuste aos dados enquanto que o método Elastic Net demonstrou maior poder de predição. O método LASSO adaptativo, quando comparado ao LASSO, apresentou bom ajuste aos dados, no entanto, sua capacidade de prever os resultados não teve boa performance. Fan et al. (2015) destacam que, para este trabalho, foi usado uma quantidade pequena de variáveis e recomendam a análise com mais variáveis para analisar a performance das técnicas.

Para a seleção de variáveis também são comumente empregados os métodos *Stepwise*, que utilizam testes repetitivos de significância para incluir ou retirar regressores do modelo.

2.3.4 Métodos *Stepwise*

Em casos reais geralmente existem diversos regressores candidatos, dos quais pretende-se selecionar apenas os que apresentem contribuições significativas ao modelo. Uma das possibilidades para se obter as variáveis dependentes significativas são os Métodos de Regressão *Stepwise*, que podem ser divididos em três tipos principais: Inclusão Passo a Frente (*Forward*), Eliminação Passo Atrás (*Backward*) e Seleção Passo-a-Passo (*Stepwise*) (MONTGOMERY; PECK; VINING, 2012). Para a definição do modelos de DEC e FEC, foi utilizada a seleção *Stepwise* (ANEEL, 2014).

A Inclusão Passo a Frente fundamenta-se a na premissa de que não há nenhuma variável regressora no modelo além do intercepto. A partir disto, inclui-se a variável com maior significância no modelo, que será a que possui maior correlação simples com a variável resposta no modelo. Após, analisa-se a contribuição de cada regressor candidato ao modelo e, caso a contribuição seja maior que o valor de corte definido pelo analista, a variável é incluída no modelo. Este processo é repetido até que não haja mais regressores a serem incluídos no modelo (MONTGOMERY; PECK; VINING, 2012).

A Eliminação Passo Atrás inicia-se com a inclusão de todas as variáveis candidatas no modelo e, a cada passo, elimina-se a variável com menor contribuição ao modelo. Posteriormente, analisa-se cada atributo considerando que tenha sido a última variável a ser incluída no modelo. O atributo que apresentar valor de significância, em relação ao modelo, menor que o valor mínimo estabelecido na análise, é eliminado. Este processo é repetido até que todos os atributos com significância menor que o limite tenham sido eliminados (MONTGOMERY; PECK; VINING, 2012).

A Regressão *Stepwise* é uma combinação dos dois métodos citados anteriormente, que se inicia da mesma forma que a Inclusão Passo a Frente, isto é, o modelo inicia-se com a variável que possui maior significância em relação à variável resposta. Posteriormente, adiciona-se o regressor cuja significância é maior que a significância mínima para

entrar no modelo e, a cada passo consecutivo, a significância de todas as variáveis incluídas são reavaliadas. Na Regressão *Stepwise* é possível que um regressor incluído anteriormente no modelo seja eliminado nos passos subsequentes caso o valor de sua significância em relação ao modelo torne-se inferior ao limite com a inclusão de outros regressores (MONTGOMERY; PECK; VINING, 2012). Em outras palavras, um regressor pode perder significância no modelo na medida em que outros regressores são adicionados ao modelo.

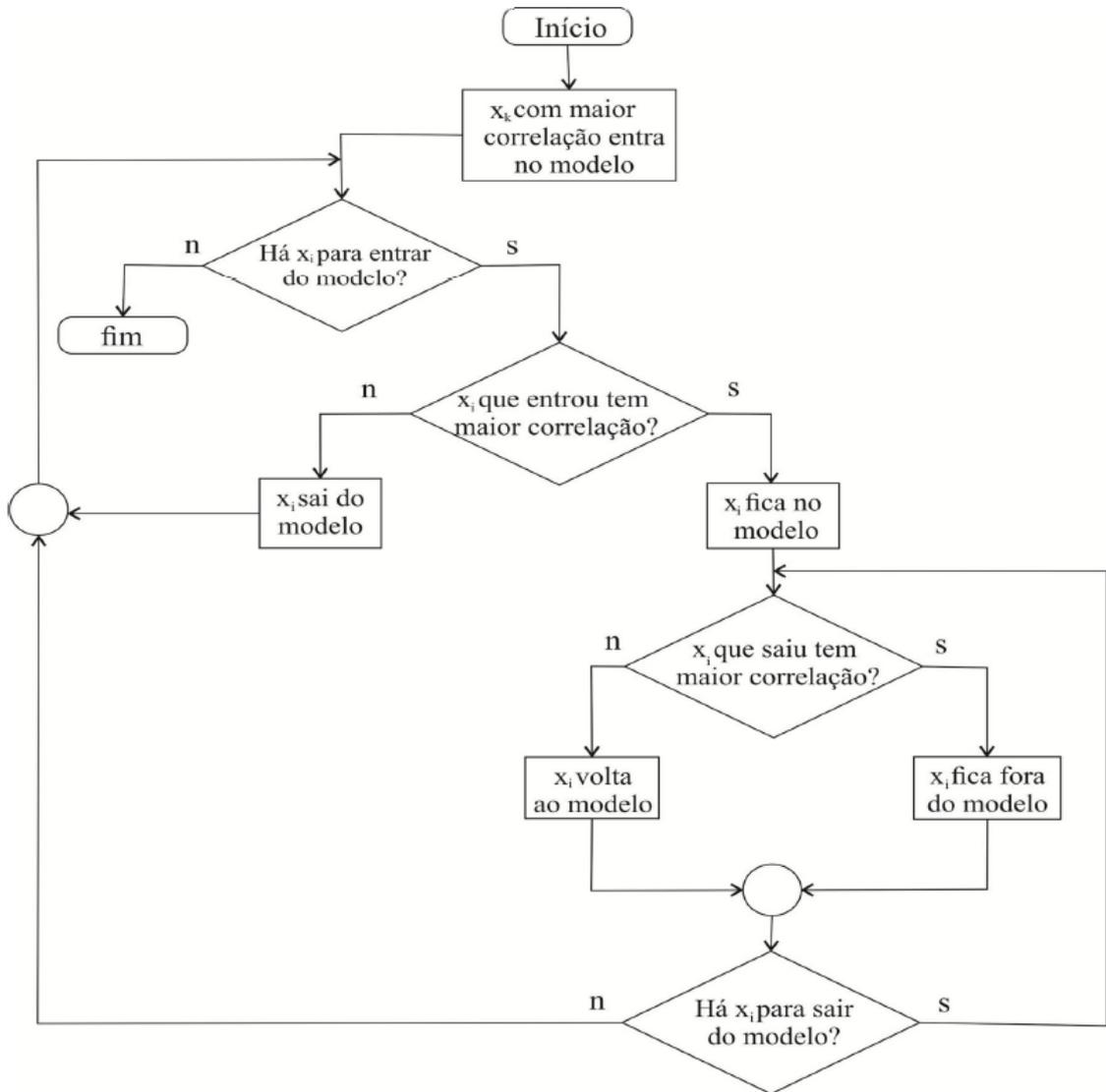
Tanto a Inclusão Passo a Frente quanto a Eliminação Passo Atrás avaliam apenas a retirada ou a inserção de variáveis no modelo, respectivamente. Uma vez que a Regressão *Stepwise* realiza os dois procedimentos citados anteriormente, o modelo final obtido tem chances de ser o melhor se comparado aos outros métodos *Stepwise* (RAWLINGS; PANTULA; DICKEY, 2011).

Todavia, a Regressão *Stepwise* não garante que o modelo final é o melhor para descrever a variável resposta, o que pode levar a interpretação que este é o modelo final, quando de fato existem outros modelos tão adequados quanto este (MONTGOMERY; PECK; VINING, 2012).

Além disso, a Regressão *Stepwise* apresenta problemas quanto à replicabilidade, visto que uma dada variável incluída no modelo pode apresentar significância infinitesimalmente maior que outras devido à erros de amostragem, dessa forma, as escolhas subsequentes serão afetadas por essa exclusão (THOMPSON, 1995).

A Figura 2.7 apresenta o fluxograma dos procedimentos de uma Regressão *Stepwise* onde no primeiro processo é inserido o regressor com maior correlação e, após, é verificado se há variáveis a serem incluídas e também a verificação da significância das variáveis já existentes no modelo.

Figura 2.7 – Fluxograma dos Procedimentos do Método *Stepwise*.



Fonte: Alves, Lotufo e Lopes (2013).

No trabalho de Smith (2018), a simulação de Monte Carlo é utilizada para demonstrar que a regressão *Stepwise* pode incluir variáveis indesejáveis no modelo em detrimento de variáveis causais. Em sua conclusão, Smith (2018) afirma que as desvantagens da regressão *Stepwise* são mais evidentes quando há um elevado número de variáveis preditoras. Além disso, é destacado que a regressão *Stepwise* não tem capacidade de prever variáveis com precisão.

A qualidade de um modelo pode ser avaliada de diversas formas, como, por exemplo, a qualidade do ajuste. Para tanto, podem ser utilizados os indicadores coeficiente de determinação e coeficiente de determinação ajustado.

2.3.5 Coeficiente de Determinação e Coeficiente de Determinação Ajustado

Existem diversos indicadores que avaliam a qualidade do ajuste do modelo de regressão aos dados, dentre eles está o coeficiente de determinação, ou R^2 . Esta métrica indica, em porcentagem, o quanto as variações na variável resposta são explicadas pelo modelo. Para tanto, analisa-se a proporção entre a SQR e a variação do modelo em relação à média das observações da variável dependente, também chamada de soma total de quadrados (STQ) (GUJARATI; PORTER, 2011). Logo, a representação matemática do R^2 é dada pela Equação 2.18.

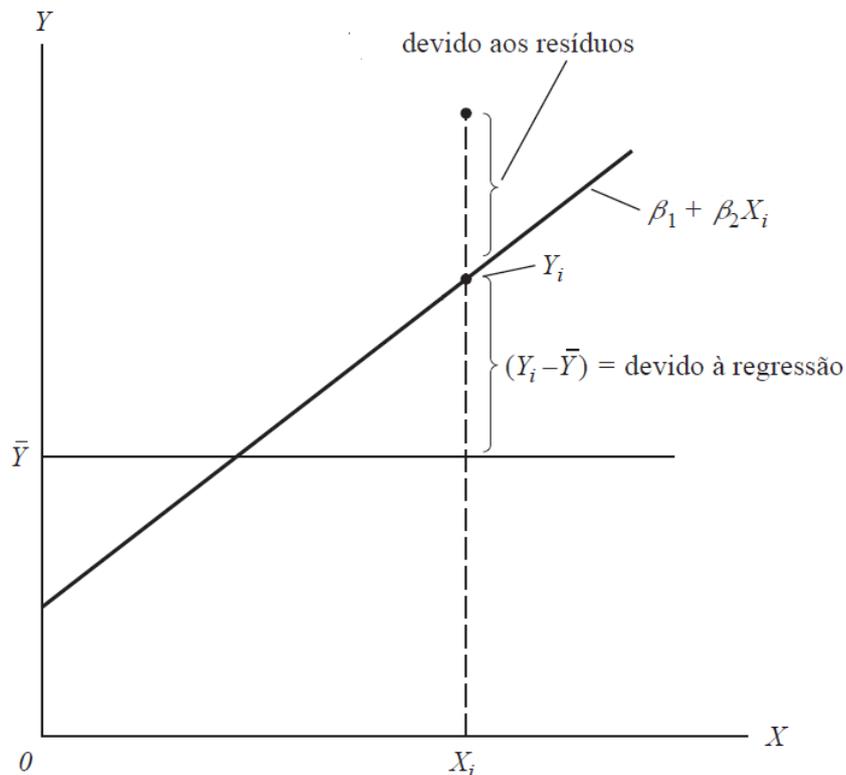
$$R^2 = \frac{SQR}{STQ} = \frac{\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.18)$$

Em que:

- \bar{y} : média de todas as observações da variável dependente.

A Figura 2.8 representada graficamente a determinação do R^2 , onde é possível observar o processo de obtenção dos resíduos e a diferença entre o modelo e o valor médio das observações da resposta.

Figura 2.8 – Representação Gráfica do Cálculo do R^2 .



O R^2 sempre aumenta quando uma variável é incluída no modelo, entretanto, não significa que o modelo tornou-se mais explicativo. A menos que a SQR seja menor que a original ao incluir a variável, a média do quadrado do erros será maior (MONTGOMERY; PECK; VINING, 2012). Portanto, um alto valor de R^2 implica que o modelo se ajusta bem aos dados, no entanto, não significa que tenha precisão de previsão.

De modo a contornar as desvantagens do R^2 , aplica-se um ajuste à este indicador, chamado de coeficiente de determinação ajustado, ou R_a^2 , que penaliza o R^2 ao incluir mais variáveis no modelo (MONTGOMERY; PECK; VINING, 2012), representado pela Equação 2.19.

$$R_a^2 = \frac{n-1}{(n-1)-p} \frac{\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.19)$$

Em que:

- p: quantidade de variáveis independentes;
- n: número de amostras.

Ao analisar Equação 2.19, percebe-se que caso a quantidade de variáveis aumente, o R_a^2 pode diminuir. Assim, o R_a^2 além de avaliar a qualidade do ajuste aos dados, estima o poder de predição do modelo (MONTGOMERY; PECK; VINING, 2012). Além disso, a magnitude deste indicador tende a estabilizar a partir de uma dada quantidade de variáveis regressoras (RAWLINGS; PANTULA; DICKEY, 2011).

Além de avaliar o poder explicativo de um modelo, também é pertinente analisar as variáveis deste modelo quanto a multicolinearidade, ou seja, a presença de atributos redundantes. Tal avaliação é possível por meio métrica fator de inflação de variância.

2.3.6 Fator de Inflação de Variância

Um dos problemas apresentados em modelos de regressões é a multicolinearidade, que consiste na presença de duas ou mais variáveis altamente correlacionadas entre si no modelo. Uma das formas de identificar este problema é através do fator de inflação de variância, ou em inglês *Variance Inflation Factor* (VIF). Este indicador é determinado para cada variável independente a partir do coeficiente de determinação da regressão da variável de interesse com as demais variáveis (RAWLINGS; PANTULA; DICKEY, 2011). A forma matemática para se obter o VIF é dada pela Equação 2.20.

$$VIF_j = \frac{1}{1 - R_j^2} \quad (2.20)$$

Em que:

VIF_j : VIF da variável j ;

R_j^2 : coeficiente de determinação da regressão da variável j com as demais variáveis.

Analisando a Equação 2.20, percebe-se que quanto maior for o valor de R_j^2 , maior será o valor do VIF. Em outras palavras, na medida que uma variável tiver a sua explicação por outra variável mais próxima de 100%, maior será o VIF do atributo em análise.

De acordo com Montgomery, Peck e Vining (2012), VIFs maiores que 10 implicam em problemas consideráveis com multicolinearidade. Portanto, o ideal é que as variáveis de um modelo possuam, no máximo, VIF igual 10, assim como foi feito na metodologia da ANEEL, onde adotou-se o limite do VIF igual a 5.

2.3.7 Considerações Finais

Neste capítulo foi realizada uma revisão bibliográfica dos indicadores de continuidade do serviço, com enfoque no modelo atual utilizado pela ANEEL para predição dos indicadores DEC e FEC. Foram também detalhados alguns aparatos estatísticos para definição de modelos de previsão de variáveis, visando apresentar alternativas à técnica empregada pela ANEEL.

Foi possível observar que as ferramentas estatísticas apresentadas dispõem de características que solucionam os problemas apresentados pela regressão *Stepwise* aplicada pela ANEEL, como capacidade de previsão de variáveis e replicabilidade.

3 METODOLOGIA

Os indicadores de continuidade do serviço são essenciais para avaliar as distribuidoras quanto a continuidade do fornecimento de energia elétrica. Estes indicadores estão sujeitos tanto a aspectos que são gerenciáveis, como a capacidade da distribuidora de reestabelecer o fornecimento de energia, quanto a aspectos que não são gerenciáveis, como incidência de descargas atmosféricas. Uma vez que o limite destes indicadores é estabelecido por meio de uma análise comparativa, a ANEEL realiza a comparação das distribuidoras considerando apenas aspectos não gerenciáveis, partindo do pressuposto que conjuntos semelhantes devem possuir desempenho similar.

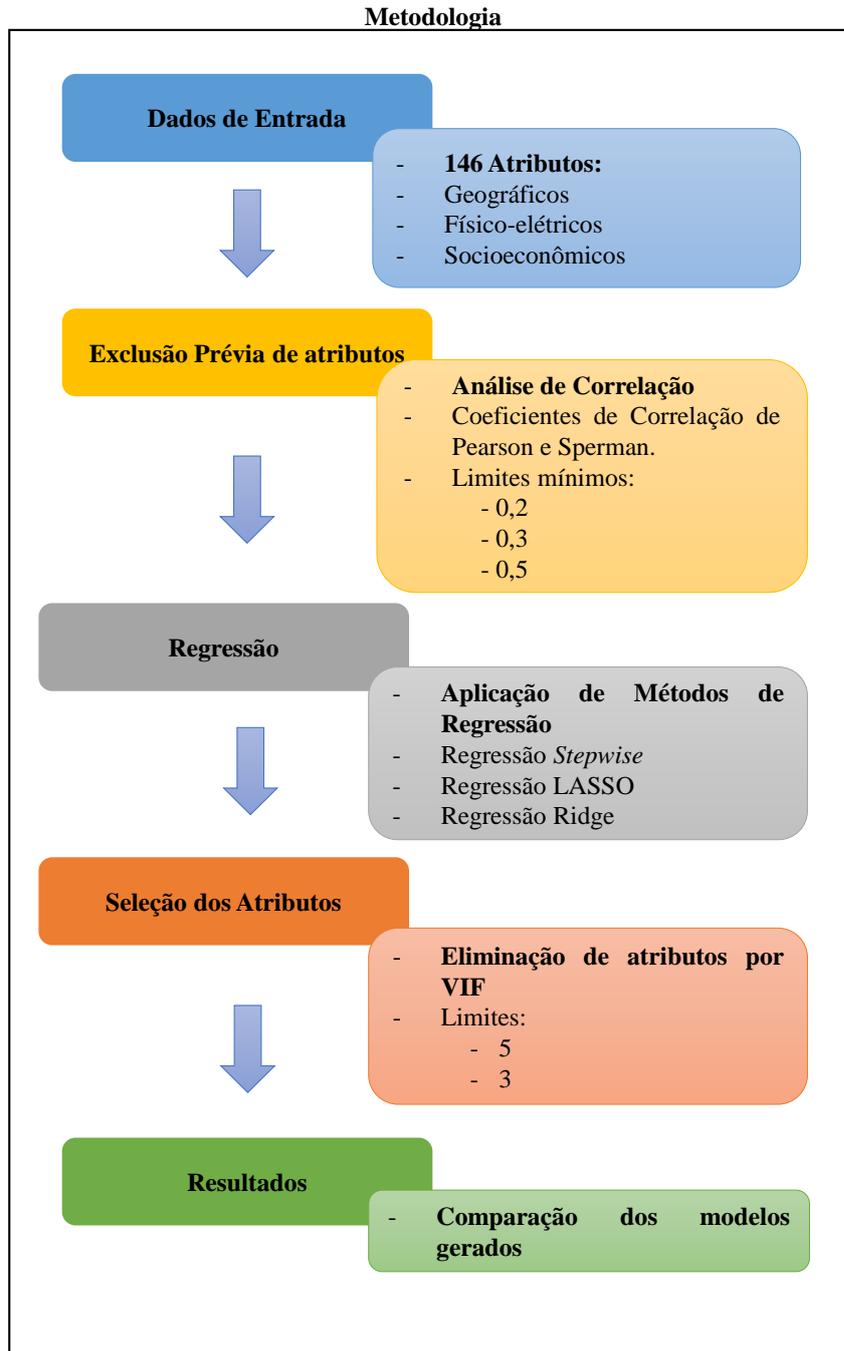
A metodologia para definição dos limites de continuidade recebeu a última atualização em 2014. Dentre as mudanças relativas aos atributos que caracterizam os conjuntos, está o emprego de uma ferramenta matemática para conceber os modelos estatísticos dos indicadores de continuidade coletivos.

É notório que um modelo teórico, por mais robusto que seja, não será capaz de representar verdadeiramente as características elétricas e geográficas das distribuidoras brasileiras, em razão da alta heterogeneidade que há entre os conjuntos de unidades consumidoras presentes no Brasil. Desta forma, deve haver uma busca constante por melhorias na metodologia de definição dos limites norteada por contribuições colhidas em audiências e consultas públicas.

Posto isto, o presente trabalho tem como objetivo analisar a relevância estatística de diferentes modelos para definição dos limites de continuidade. A metodologia proposta baseia-se nos mesmos procedimentos utilizados pela ANEEL, entretanto, há a substituição das técnicas estatísticas empregadas, visando construir e avaliar estudos de caso. A estrutura da metodologia proposta é mostrada na Figura 3.1.

Nesta análise, considera-se que a apuração dos atributos utilizados foi realizada corretamente, visto que estes dados são a base de toda a análise a ser feita. A apuração destas informações está diretamente relacionada com a qualidade dos modelos a serem gerados, uma vez que estes registros serão utilizados para modelar as variáveis resposta DEC e FEC.

Figura 3.1 – Metodologia Proposta.



Fonte: do Autor

Conforme Figura 3.1, a metodologia proposta é dividida em 5 etapas. A primeira etapa compreende os dados de entrada, os quais são os 146 atributos presentes no apêndice 3 dos resultados da Audiência Pública nº029/2014, no qual estão os atributos apurados para todos os conjuntos brasileiros do ano de 2013, bem como os valores apurados de DEC e FEC dos anos de 2011, 2012 e 2013. No Anexo A constam o nome e a sigla destas variáveis.

Destes 146 atributos, 89 são oriundos da BDGD, isto é, são dados fornecidos pelas distribuidoras relativos aos dados técnicos do sistema de distribuição, às informações comerciais e aos dados físico-contábeis da base de ativos. As informações socioeconômicas são sintetizadas por 13 atributos provenientes de outras bases de dados. As informações de vegetação são dadas por 20 atributos que apresentam informações de altura da vegetação. Os dados relacionados a descargas atmosféricas e precipitação são caracterizados por 3 atributos e os dados de arruamento por 21 atributos.

Estas variáveis presentes no apêndice 3 dos resultados da Audiência Pública nº 029/2014 foram utilizadas para a definição do modelo vigente empregado pela ANEEL, desta forma, a contraposição dos modelos obtidos nos estudos de caso com o modelo da ANEEL torna-se mais coerente.

Após a definição dos dados de entrada, na segunda etapa é realizada a exclusão prévia, fundamentada na aplicação de coeficientes de correlação, onde são eliminados os atributos com alta correlação entre si, optando-se pelos atributos com maior correlação com DEC e FEC. Além disso, também são retirados os atributos com baixa correlação com DEC e FEC, aplicando-se três limites mínimos de correlação, que são, em módulo: 0,2, 0,3 e 0,5. Optou-se apenas por estes limites de correlação pois poucos atributos apresentam correlação superiores a 0,5. Portanto, poucos atributos remanesceriam da exclusão por correlação caso fossem estipulados limites superiores a esses.

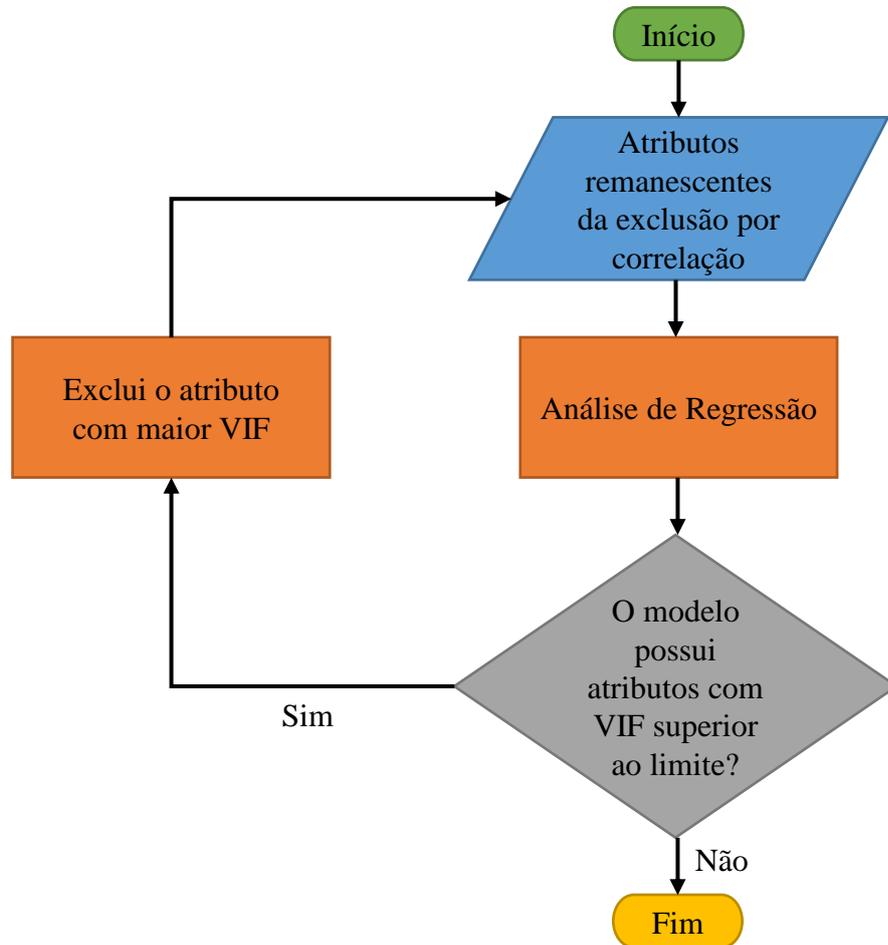
Esta filtragem prévia de atributos permite a redução de variáveis para a próxima etapa. Objetivando a normalização dos dados, os atributos remanescentes da eliminação por correlação são normalizados de acordo com a Equação 2.7.

Tendo em posse os atributos candidatos ao modelo após a filtragem prévia, a próxima etapa, de acordo com a Figura 3.1, consiste em realizar a regressão desta variáveis. As técnicas de regressão utilizadas para construir os modelos são: Regressão *Stepwise* com significância de entrada igual a 1%, Regressão LASSO e Regressão Ridge. Assim como na análise dos atributos realizada pela ANEEL, em cada regressão serão gerados dois modelos para cada indicador, sendo que um dos modelos contempla as variáveis socioeconômicas e o outro desconsidera estas variáveis.

Após a concepção dos modelos, faz-se necessário eliminar a multicolinearidade que há entre as variáveis presentes no modelo. Para tanto, os atributos com maior grau de multicolinearidade serão eliminados de acordo com o VIF. Nesta etapa da análise, são utilizados dois limites máximos do VIF para manter as variáveis no modelo: 5 e 3.

A Figura 3.2 apresenta o fluxograma de definição do modelo, que contempla as etapas de regressão e seleção de atributos.

Figura 3.2 – Definição do modelo.



Fonte: do Autor

Por fim, os modelos gerados com base nos parâmetros anteriormente citados são comparados quanto a quantidade de variáveis que os compõem e quanto ao poder explicativo de cada modelo. Além disso, realiza-se a avaliação da variação do poder explicativo por modelo e por variável presente nos modelos tomando como referência o modelo construído ao replicar as etapas da definição dos atributos pela ANEEL.

Para a definição dos modelos teóricos foi utilizado o *software SAS University Edition*, desenvolvido pela empresa de inteligência analítica SAS, para realizar a análise estatística dos dados.

3.1 SAS UNIVERSITY EDITION

O *SAS University Edition* fornece *procedures*, que são funções utilizadas para análise e manipulação de dados onde é necessário apenas incluir os argumentos, como variáveis de entrada e diagnósticos desejados. Para realizar as análises propostas neste

trabalho, são necessários três *procedures* do SAS *University Edition*: *glmselect*, *reg* e *import*.

O *procedure import* é sempre o primeiro a ser utilizado no código, pois através desse procedimento os dados fornecidos pela ANEEL em formato .XLSX são lidos e escritos em um conjunto de dados SAS.

Nos casos onde será empregada a regressão LASSO é utilizado o *procedure glmselect*, que permite realizar apenas procedimentos de seleção de variáveis a partir do conjunto de dados gerado pelo *procedure import*.

Nos casos onde serão empregadas as regressões *Stepwise* e *Ridge* utiliza-se o *procedure reg*, que, dentre suas funcionalidades, permite realizar técnicas de regressão e apresentar diagnósticos de ajuste de modelos. Desta forma, os dados escritos pelo *procedure import* serão utilizados para aplicar essas técnicas de regressão.

Ambos os *procedures glmselect* e *reg* dispõem do diagnóstico de VIF dos modelos. Assim sendo, as variáveis que apresentarem VIF maior que o limite da análise serão retiradas do argumento dos *procedures* até que o modelo possua apenas atributos com VIF inferior ao limite.

3.2 CONSIDERAÇÕES FINAIS

Neste capítulo foi apresentada a estrutura da metodologia proposta para realizar a análise estatística para seleção dos atributos que caracterizam os indicadores coletivos. Foram detalhadas as cinco etapas que compõem a análise proposta. Além disso, foi apresentado o fluxograma do processo de definição dos modelos estatísticos, que compreende as etapas de regressão e seleção dos atributos, além disso, este fluxograma é válido para as três regressões propostas neste trabalho.

Por fim, discorreu-se sobre a forma como as análises estatísticas serão efetuadas no *software SAS University Edition*.

O próximo capítulo apresenta os estudos de caso formulados, assim como os resultados obtidos e a comparação dos modelos gerados.

4 ESTUDO DE CASO

O presente capítulo apresenta a aplicação da metodologia proposta de análise estatística para seleção dos atributos que caracterizam os indicadores de continuidade, apresentada no Capítulo 3. Para validação desta metodologia, são propostos 8 estudos de caso, variando-se o critério de correlação para selecionar as variáveis que devem seguir para próxima etapa. Além disso, são utilizadas três técnicas estatísticas de regressão. Por fim, eliminam-se as variáveis dos modelos com base no VIF.

Tanto as análises de correlação quanto as de regressão foram realizadas utilizando o *software SAS University Edition*. Este *software* permite realizar o processamentos e a análise estatística de dados (SAS, 2021).

Como dados de entrada, foram utilizados os registros dos 146 atributos presentes no apêndice III da Nota Técnica nº102, que, após a análise de correlação, foram normalizados de acordo com a equação 2.7. O nome, a unidade e a sigla destes atributos são apresentados no Anexo A.

4.1 ESTUDO DE CASO 1

O primeiro estudo de caso consiste em replicar o processo de definição dos modelos realizado pela ANEEL, onde para cada indicador, foram desenvolvidos dois modelos: considerando e desconsiderando as variáveis socioeconômicas. Para tanto, os 146 atributos definidos pela ANEEL foram submetidos a uma exclusão prévia, visando eliminar os atributos que possuem correlação de Pearson ou Spearman com DEC ou FEC inferior a 0,2, conforme mostra a Tabela 4.1.

Tabela 4.1 – Parâmetros utilizados no estudo de caso 1.

Parâmetros do Estudo de Caso 1	
Limite mínimo de correlação	0,2
Limite máximo de VIF	5
Técnica de Regressão	<i>Stepwise</i>

Fonte: do Autor.

4.1.1 DEC

Para o modelo do DEC considerando as variáveis socioeconômicas e desconsiderando as variáveis socioeconômicas, remanesceram 69 e 58 variáveis independentes, res-

pectivamente. Estas variáveis foram submetidas a seleção de atributos visando eliminar a multicolinearidade, onde aplicou-se a regressão *Stepwise* eliminando a cada passo a variável com maior VIF até todos os atributos apresentarem VIF inferior a 5. A seleção resultou em 20 variáveis considerando as socioeconômicas e 10 variáveis desconsiderando-as.

Os atributos oriundos da seleção foram novamente submetidas a regressão *Stepwise*, resultando em 17 variáveis independentes para o DEC com atributos socioeconômicos e 9 atributos para o DEC sem atributos socioeconômicos. Este passo difere-se do anterior devido à remoção no *script* de atributos excluídos durante a análise de regressão *Stepwise*. A Tabela 4.2 apresenta os resultados da regressão *Stepwise* para os modelos do DEC.

Tabela 4.2 – Resultado da regressão *Stepwise* para o modelo do DEC considerando e desconsiderando variáveis socioeconômicas.

Com Variáveis Socioeconômicas				Sem Variáveis Socioeconômicas		
N	Atributo	R ²	R ² Modelo	Atributo	R ²	R ² Modelo
1	PAREDE_IN	0,3255	0,3255	PC_NUC_URB	0,3249	0,3249
2	PC_NUC_URB	0,1634	0,4889	PC_VRAM	0,0905	0,4153
3	PC_ERMT_3F	0,0492	0,5381	PC_ERMT_3F	0,0590	0,4743
4	PLUV	0,0432	0,5813	PLUV	0,0462	0,5205
5	PC_NUC_COM	0,0264	0,6077	CM_NUC_RES	0,0405	0,5610
6	PC_VRAM	0,0222	0,6298	NUC_IND	0,0252	0,5862
7	CM_NUC_RES	0,0115	0,6414	PC_NUC_COM	0,0122	0,5985
8	PC_NUC_RUR	0,0094	0,6507	NUC_RUR	0,0080	0,6065
9	IDHM	0,0089	0,6596	ERMT_AREA	0,0097	0,6162
10	PC_ENE_RUR	0,0066	0,6662			
11	VIO	0,0081	0,6742			
12	TD_ERMT	0,0046	0,6788			
13	LIXO	0,0026	0,6814			
14	PEXP	0,0082	0,6896			
15	ERMT_URB	0,0012	0,6908			
16	ENE_IND	0,0028	0,6936			
17	PNI_NURB	0,0009	0,6945			

Fonte: do Autor.

Observando a Tabela 4.2, nota-se que para ambos os modelos, o poder explicativo do modelo torna-se menos sensível à inclusão de novas variáveis a partir do sétimo atributo incluído, cujo poder explicativo adicional é inferior a 2%. Além disso, a inclusão de novas variáveis aumenta a susceptibilidade do modelo ao erro amostral. Desta forma, os seis primeiros atributos são os mais relevantes para os modelos do DEC, conforme é apresentado na Tabela 4.3.

Tabela 4.3 – Modelos Selecionados para o DEC.

Com Varáveis Socioeconômicas		Sem Varáveis Socioeconômicas	
Atributos	R ²	Atributos	R ²
PAREDE_IN	0,6298	PC_NUC_URB	0,5862
PC_NUC_URB		PC_VRAM	
PC_ERMT_3F		PC_ERMT_3F	
PLUV		PLUV	
PC_NUC_COM		CM_NUC_RES	
PC_VRAM		NUC_IND	

Fonte: do Autor.

Nota-se que as variáveis selecionadas para o modelo do DEC sem atributos socioeconômicos são idênticas às variáveis utilizados no modelo da ANEEL, apresentado no Quadro 2.1. Assim como foi definido os modelos do DEC, deve-se definir os modelos do FEC.

4.1.2 FEC

A exclusão prévia de atributos em relação ao FEC considerando e desconsiderando as variáveis socioeconômicas resultou em 67 e 56 variáveis independentes, respectivamente. De forma análoga ao processo adotado para os modelos de DEC, procedeu-se a seleção destes atributos remanescentes da exclusão prévia, no sentido de eliminar a multicolinearidade. A seleção resultou em 15 variáveis considerando as socioeconômicas e 12 variáveis desconsiderando-as.

Tendo em posse os atributos resultantes da seleção, procedeu-se a regressão *Stepwise*, que definiu um modelo com 15 variáveis independentes para o FEC com atributos socioeconômicos e 10 atributos para o FEC sem atributos socioeconômicos, conforme mostra a Tabela 4.4.

Tabela 4.4 – Resultado da regressão *Stepwise* para o modelo do FEC considerando e desconsiderando variáveis socioeconômicas.

Com Variáveis Socioeconômicas				Sem Variáveis Socioeconômicas		
N	Atributo	R ²	R ² Modelo	Atributo	R ²	R ² Modelo
1	LIXO	0,3563	0,3563	TD_ERMT	0,2218	0,2218
2	PLUV	0,0957	0,4520	PLUV	0,1484	0,3702
3	ERMT_AREA	0,0825	0,5345	PC_VRAM	0,0984	0,4686
4	PEXP	0,0408	0,5753	PC_NUC_URB	0,0451	0,5137
5	NUC_COM	0,0350	0,6103	CM_NUC	0,0207	0,5343
6	ERMT_1F	0,0184	0,6287	PC_NUC_RES	0,0137	0,5480
7	PC_VRAM	0,0202	0,6489	PC_ERMT_3F	0,0113	0,5593
8	TD_1F	0,0204	0,6694	NUC_RUR	0,0039	0,5632
9	PAREDE_IN	0,0079	0,6773	CM_NUC_RES	0,0016	0,5648
10	IDHM	0,0048	0,6821	DENS_EST_PAV	0,0018	0,5666
11	ALIM	0,0046	0,6867			
12	CM_NUC	0,0036	0,6903			
13	ENE_IND	0,0042	0,6945			
14	PC_NUC_URB	0,0011	0,6955			
15	PC_EST_ARR	0,0013	0,6968			

Fonte: do Autor.

Verifica-se que para os dois modelos do FEC, a partir da inclusão da sexta variável o poder explicativo adicional é menor que 2%. Desta forma, pode-se definir que o modelo com 5 variáveis considerando e desconsiderando as variáveis socioeconômicas já apresenta poder explicativo suficiente para caracterizar o indicador, como mostra a Tabela 4.5.

Tabela 4.5 – Modelos Selecionados para o FEC.

Com Varáveis Socioeconômicas		Sem Varáveis Socioeconômicas	
Atributos	R ²	Atributos	R ²
LIXO	0,6103	TD_ERMT	0,5343
PLUV		PLUV	
ERMT_AREA		PC_VRAM	
PEXP		PC_NUC_URB	
NUC_COM		CM_NUC	

Fonte: do Autor.

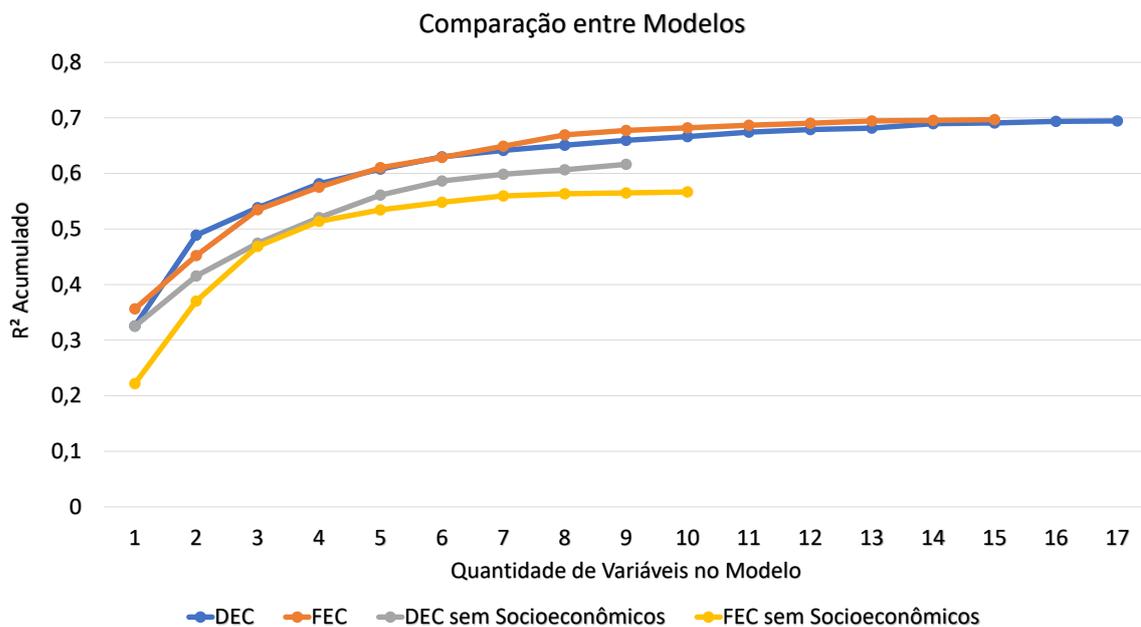
Comparando o modelo sem variáveis socioeconômicas selecionado com o modelo da ANEEL, apresentado no Quadro 2.1, nota-se que há uma variável a menos e, das 5 variáveis selecionados, apenas as variáveis PLUV, PC_VRAM, e PC_NUC_URB são

idênticas as do modelo da ANEEL. Essas alterações nas variáveis que compõem o modelo causam mudanças na trajetória dos limites dos conjuntos, visto que os limites são definidos por meio da clusterização e esta, por sua vez, é realizada por com base nos atributos que fazem parte do modelo do indicador.

4.1.3 Resultados do Estudo de Caso 1

Analisando os resultados obtidos ao reproduzir a análise estatística adotada pela ANEEL, percebe-se que tanto para o FEC quanto para o DEC, os modelos com variáveis socioeconômicas possuem maior poder explicativo, como mostra a Figura 4.1, o que demonstra a relevância das informações socioeconômicas para os modelos estatísticos. Apesar de a ANEEL obter os mesmos resultados, optou por utilizar o modelo sem informações socioeconômicas devido à periodicidade de atualização destes dados.

Figura 4.1 – Modelos do Caso 1



Fonte: do Autor.

Em relação às variáveis selecionadas, apenas o modelo do DEC sem variáveis socioeconômicas foi congruente com o obtido pela ANEEL. Já o modelo selecionado para o FEC sem variáveis socioeconômicas possui 5 atributos, enquanto que o definido pela ANEEL possui 6 atributos, sendo que apenas 3 variáveis são comuns a ambos. Os demais modelos considerando os dados socioeconômicos também diferiram da ANEEL. Estas diferenças, como apresentado durante a revisão bibliográfica, podem ser causadas pela baixa replicabilidade inerente a regressão *Stepwise*. Estas mudanças, se aplicadas no método

de clusterização dinâmico, podem modificar os limites de continuidade das distribuidoras, uma vez que o agrupamento de conjuntos é baseia-se nos atributos utilizados para definir a semelhança entre conjuntos.

O próximo estudo de caso aborda a utilização dos mesmo parâmetros do presente estudo, variando-se apenas o limite do VIF.

4.2 ESTUDO DE CASO 2

No segundo estudo de caso utilizou-se o critério de correlação inferior a 20% para eliminação dos atributos e a regressão *Stepwise* para gerar os modelos. Para a eliminação de multicolinearidade foi definido o VIF inferior a 3 para manter as variáveis no modelo. Estes parâmetros são apresentados na Tabela 4.6.

Tabela 4.6 – Parâmetros utilizados no estudo de caso 2.

Parâmetros do Estudo de Caso 2	
Limite mínimo de correlação	0,2
Limite máximo de VIF	3
Técnica de Regressão	<i>Stepwise</i>

Fonte: do Autor.

Desta forma, a Tabela 4.7 apresenta os modelos obtidos para os indicadores DEC e FEC considerando e desconsiderando as informações socioeconômicas com seus respectivos atributos e coeficientes de determinação.

Tabela 4.7 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 2.

Com Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R²	Atributos	R²
LIXO	0,6148	LIXO	0,6067
PC_NUC_URB		PLUV	
IDHM		PC_VRAM	
PLUV		NUC_COM	
PC_NUC_COM		PC_ERMT_3F	
Sem Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R²	Atributos	R²
PC_NUC_URB	0,5610	PC_VRAM	0,5219
PC_VRAM		PLUV	
PC_ERMT_3F		PC_ENE_COM	
PLUV		ENE_PNI	
CM_NUC_RES		NUC_IND	

Fonte: do Autor.

Comparando os modelos do DEC obtidos neste estudo de caso com os modelos do estudo de caso 1, percebe-se que, ao reduzir o VIF limite para 3, houve um decréscimo na quantidade de atributos e no poder explicativo dos modelos. Além disso, o modelo com dados socioeconômicos é o que possui maior poder explicativo.

Quanto aos modelos do FEC, verifica-se que não ocorreu alterações na quantidade de atributos, no entanto, o poder explicativo reduziu para ambos os modelos do FEC.

O próximo estudo de caso apresenta a utilização da Regressão *Stepwise* aumentando o critério de correlação.

4.3 ESTUDO DE CASO 3

Para o terceiro estudo de caso foi adotado uma de correlação mínima de 30% para manter os atributos na análise. A técnica de regressão utilizada pra gerar os modelos foi a Regressão *Stepwise* e, para a eliminação da multicolinearidade, adotou-se o VIF inferior a 5 como limite para os atributos serem mantidos no modelo, conforme apresenta a Tabela 4.8.

Tabela 4.8 – Parâmetros utilizados no estudo de caso 3.

Parâmetros do Estudo de Caso 3	
Limite mínimo de correlação	0,3
Limite máximo de VIF	5
Técnica de Regressão	<i>Stepwise</i>

Fonte: do Autor.

Os modelos gerados, considerando e desconsiderando as variáveis socioeconômicas, a partir destes parâmetros são apresentados na Tabela 4.9.

Tabela 4.9 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 3.

Com Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R ²	Atributos	R ²
PAREDE_IN	0,5667	TD_ERMT	0,3323
PC_NUC_URB		PC_ERMT_1F	
PC_ERMT_1F		NUC_URB	
IDHM			
Sem Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R ²	Atributos	R ²
PC_NUC_URB	0,5039	ERMT_AREA	0,3792
PC_VRAM		TD_URB	
PC_ERMT_3F		PC_ERMT_1F	
NUC_IND			

Fonte: do Autor.

Comparando os modelos deste estudo de caso com os do primeiro estudo de caso, que diferencia-se deste apenas pelo critério de correlação, é possível notar que para o indicador DEC houve uma redução de dois atributos no modelo e um decréscimo do poder explicativo. Assim como nos estudos de caso anteriores, o modelo do DEC com variáveis socioeconômicas possui o maior poder explicativo.

Os modelos do indicador FEC, se comparados aos do primeiro estudo de caso, possuem três variáveis a menos e, além disso, o poder explicativo dos modelos gerados neste estudo de caso também são inferiores. Diferentemente dos modelos obtidos anteriormente, neste estudo de caso o modelo do FEC sem variáveis econômicas é o que possui maior poder explicativo.

O próximo estudo de caso retrata a utilização dos mesmos critérios empregados no presente estudo de caso, no entanto, o limite máximo do VIF é reduzido.

4.4 ESTUDO DE CASO 4

No estudo de caso 4 foram eliminados da análise os atributos que possuem correlação inferior a 30% com os indicadores DEC e FEC. Os modelos foram estabelecidos por meio da Regressão *Stepwise* e o VIF limite definido para manter os atributos nos modelos foi 3. Estes critérios são apresentados na Tabela 4.10.

Tabela 4.10 – Parâmetros utilizados no estudo de caso 4.

Parâmetros do Estudo de Caso 4	
Limite mínimo de correlação	0,3
Limite máximo de VIF	3
Técnica de Regressão	<i>Stepwise</i>

Fonte: do Autor.

Os resultados obtidos ao adotar estes critérios são apresentados na Tabela 4.11.

Tabela 4.11 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 4.

Com Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R²	Atributos	R²
PAREDE_IN	0,5283	NUC_URB	0,3717
PC_ERMT_1F		PC_ERMT_1F	
NUC_COM		CM_NUC	
		DENS_EST_PAV	
Sem Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R²	Atributos	R²
PNI_TD	0,4593	ERMT_AREA	0,3643
PC_VRAM		PC_ERMT_1F	
TD_3F		NUC_URB	
ENE_PNI			

Fonte: do Autor.

Ao contrapor os modelos do indicador DEC estabelecidos neste estudo de caso com os modelos do estudo de caso 3, cuja diferença entre estes estudos de caso é o VIF limite, verifica-se que, ao diminuir o VIF limite, ocorreu uma redução no poder explicativo dos modelos do DEC e, além disso, foi reduzida a quantidade de atributos no modelo com dados socioeconômicos.

Observando os modelos do indicador FEC, observa-se que, da mesma forma que ocorreu com o indicador DEC, ao reduzir o VIF limite, o poder explicativo do modelo sem variáveis socioeconômicas também reduziu, no entanto, o modelo com variáveis socioeconômicas possui um maior poder explicativo se comparado ao do terceiro estudo de caso.

O próximo estudo de caso abrange a utilização da Regressão *Stepwise* aumentando o limite de correlação.

4.5 ESTUDO DE CASO 5

No quinto estudo de caso a correlação limite adotada para os atributos prosseguirem na análise foi de 50%. Como técnica de regressão, manteve-se a Regressão *Stepwise* para gerar os modelos, onde foram eliminados os atributos com VIF maior que 5. Estes parâmetros estão sintetizados na Tabela 4.12.

Tabela 4.12 – Parâmetros utilizados no estudo de caso 5.

Parâmetros do Estudo de Caso 5	
Limite mínimo de correlação	0,5
Limite máximo de VIF	5
Técnica de Regressão	<i>Stepwise</i>

Fonte: do Autor.

Os modelos obtidos utilizando estes parâmetros são apresentados na Tabela 4.13.

Tabela 4.13 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 5.

Com Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R²	Atributos	R²
VRAM	0,5499	VRAM	0,4330
ENE_COM		PAREDE_IN	
PAREDE_IN		PC_ERMT_3F	
PC_ERMT_1F			
Sem Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R²	Atributos	R²
VRAM	0,4966	VRAM	0,3395
ENE_COM		PC_ERMT_3F	
PC_ERMT_1F			

Fonte: do Autor.

Comparando os modelos do indicador DEC estabelecidos neste estudo de caso com os modelos do estudo de caso 1, cuja diferença entre os critérios de análise é o limite de correlação, percebe-se que houve uma redução na quantidade de variáveis nos modelos, sendo duas no modelo sem atributos socioeconômicos e três no modelo com atributos socioeconômicos. Além disso, o poder explicativo dos modelos reduziu neste estudo de caso.

Analisando os modelos do FEC, nota-se que a redução de variáveis no modelo foi ainda mais acentuada, onde o modelo sem variáveis socioeconômicas passou a ser composto por três variáveis e o modelo com variáveis socioeconômicas passou a ser composto por duas variáveis. Também houve uma redução no poder explicativo dos modelos.

O próximo estudo de caso apresenta a construção dos modelos utilizando as mesmas técnicas do presente estudo com a redução do VIF limite.

4.6 ESTUDO DE CASO 6

Para o estudo de caso 6 foi mantido o limite mínimo de correlação igual a 50% para os atributos prosseguirem na análise. A técnica de regressão utilizada foi a Regressão *Stepwise*, através da qual foram gerados e, posteriormente, definidos os modelos com variáveis que apresentaram o VIF inferior a 3. A Tabela 4.14 apresenta estes parâmetros.

Tabela 4.14 – Parâmetros utilizados no estudo de caso 6.

Parâmetros do Estudo de Caso 6	
Limite mínimo de correlação	0,5
Limite máximo de VIF	3
Técnica de Regressão	<i>Stepwise</i>

Fonte: do Autor.

Os modelos concebidos são apresentados na Tabela 4.15.

Tabela 4.15 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 6.

Com Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R ²	Atributos	R ²
PC_NUC_URB	0,5507	VRAM	0,4330
PAREDE_IN		PAREDE_IN	
PC_ERMT_1F		PC_ERMT_3F	
PC_VRAM			
Sem Variáveis Socioeconômicas			
DEC		FEC	
Atributos	R ²	Atributos	R ²
PC_NUC_URB	0,5046	VRAM	0,3395
PC_VRAM		PC_ERMT_3F	
PC_ERMT_1F			
ENE			

Fonte: do Autor.

Contraopondo os modelos formados neste estudo de caso com os modelos do estudo de caso 6, que difere-se deste pelo VIF, percebe-se que no modelo do indicador DEC com variáveis socioeconômicas ocorreu um pequeno aumento no poder explicativo enquanto que a quantidade de atributos manteve-se a mesma. Quanto ao modelo sem variáveis socioeconômicas, verificou-se um aumento no poder explicativo e na quantidade de variáveis presentes no modelo.

Analisando os modelos do FEC em relação aos modelos do estudo de caso 5, nota-se que reduzir o VIF limite não causou alterações nos modelos finais.

O próximo estudo de caso apresenta a utilização da regressão LASSO para confeccionar os modelos.

4.7 ESTUDO DE CASO 7

No sétimo estudo de caso foi utilizada a Regressão LASSO para gerar os modelos para os indicadores. O limite mínimo de correlação adotado para manter os atributos nesta análise foi 20%. Em relação a multicolinearidade, foram utilizados os modelos cujo VIF das variáveis era inferior a 5. Estes parâmetros são apresentados na Tabela 4.16.

Tabela 4.16 – Parâmetros utilizados no estudo de caso 7.

Parâmetros do Estudo de Caso 7	
Limite mínimo de correlação	0,2
Limite máximo de VIF	5
Técnica de Regressão	LASSO

Fonte: do Autor.

Os modelos obtidos neste estudo de caso são apresentados na Tabela 4.17.

Tabela 4.17 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 7.

Com Variáveis Socioeconômicas				
DEC		FEC		
Atributos	R²	Atributos	R²	
VRAM	0,6241	PAREDE_IN	0,6298	
PC_NUC_URB		LIXO		
LIXO		PC_VRAM		
PAREDE_IN		PC_ERMT_3F		
ENE		TD_ERMT		
ERMT_1F		PLUV		
PLUV		DENS_EST_PAV		
PC_NUC_COM		PC_NUC_URB		
	ERMT_URB			
	PC_ENE_COM			
	ENE_PNI			
	PEXP			
	PNI_NURB			
	IDHM			
Sem Variáveis Socioeconômicas				
DEC		FEC		
Atributos	R²	Atributos	R²	
PC_NUC_URB	0,6265	ENE_RES	0,5513	
PC_ERMT_3F		TD_ERMT		
TD_ERMT		PC_VRAM		
PC_VRAM		PC_ERMT_3F		
NUC_IND		PLUV		
CM_NUC_RES		DENS_EST_PAV		
PLUV		PC_NUC_URB		
PC_NUC_COM		NUC_IND		
DENS_EST_PAV		PNI_TD		
NUC_RUR				
PC_NUC_RES				
PC_NUC_IND				
ENE_PNI				

Fonte: do Autor.

Percebe-se que os modelos gerados neste estudo de caso são compostos por uma quantidade maior de atributos, se comparados aos modelos obtidos nos estudos de caso

anteriores.

No que tange o poder explicativo do modelo, percebe-se que também aumentou em relação aos estudos de caso anteriores, sendo que apenas o modelo do DEC com socioeconômicos dos primeiro estudo de caso possui poder explicativo superior ao deste caso.

O próximo estudo de caso apresenta a utilização da regressão Ridge para confeccionar os modelos.

4.8 ESTUDO DE CASO 8

Para o oitavo estudo de caso optou-se por utilizar a técnica de regressão Ridge para conceber os modelos dos indicadores. Neste caso, utilizou-se o limite mínimo de correlação de 20% para manter os atributos na análise. Na seleção de atributos, optou-se por manter o modelo cujas variáveis apresentassem VIF menor que 5. Estes critérios são apresentados na Tabela 4.18.

Tabela 4.18 – Parâmetros utilizados no estudo de caso 8.

Parâmetros do Estudo de Caso 8	
Limite mínimo de correlação	0,2
Limite máximo de VIF	5
Técnica de Regressão	Ridge

Fonte: do Autor.

Os modelos construídos a partir destes critério são apresentados na Tabela 4.19

Tabela 4.19 – Modelos Selecionados para os indicadores DEC e FEC no Estudo de Caso 8.

Com Variáveis Socioeconômicas					
DEC			FEC		
Atributos		R ²	Atributos		R ²
DENS_EST_PAV	PC_ENE_RUR	0,7102	DENS_EST_PAV	PC_ENE_RUR	0,6778
PC_EST_ARR	PC_ERMT_3F		PC_EST_ARR	PC_ERMT_3F	
PLUV	ALIM		PLUV	ALIM	
C_VRAM	TD_1F		PC_VRAM	TD_1F	
NUC_RUR	PNI_NURB		PC_NUC_URB	PNI_NURB	
PC_NUC_RES	ENE_PNI		NUC_IND	ENE_PNI	
PC_NUC_COM	PEXP		NUC_RUR	PEXP	
PC_NUC_IND	PAREDE_IN		PC_NUC_RES	PAREDE_IN	
ENE_IND	IDHM		ENE_IND	IDHM	
CM_NUC	LIXO		CM_NUC	LIXO	
CM_NUC_RES	SUB		CM_NUC_RES	SUB	
PC_ENE_COM	VIO		PC_ENE_COM	VIO	
Sem Variáveis Socioeconômicas					
DEC			FEC		
Atributos		R ²	Atributos		R ²
DENS_EST_PAV	CM_NUC	0,6285	DENS_EST_PAV	CM_NUC	0,6103
PC_EST_ARR	CM_NUC_RES		PC_EST_ARR	CM_NUC_RES	
PLUV	PC_ENE_COM		PLUV	PC_ENE_COM	
PC_VRAM	PC_ENE_RUR		PC_VRAM	PC_ENE_RUR	
NUC_RUR	PC_ERMT_3F		PC_NUC_URB	PC_ERMT_3F	
PC_NUC_RES	ALIM		NUC_IND	ALIM	
PC_NUC_COM	TD_1F		NUC_RUR	PC_TD_1F	
PC_NUC_IND	PNI_NURB		PC_NUC_RES	PNI_NURB	
ENE_IND	ENE_PNI		ENE_IND	ENE_PNI	

Fonte: do Autor.

Observando a Tabela 4.19, nota-se que este estudo de caso é o que possui a maior quantidade de atributos. Este fato é explicado pela técnica de regressão empregada, uma vez que a Regressão Ridge não realiza a eliminação de variáveis. Além disso, dentre todos os estudos de caso, este apresenta os modelos com maior poder explicativo.

4.9 COMPARAÇÃO DOS MODELOS

Após a construção dos estudos de caso, é realizada a comparação dos modelos obtidos. Os parâmetros utilizados em todos os estudos de caso são apresentados na Tabela 4.20, destacando os que são idênticos a análise da ANEEL.

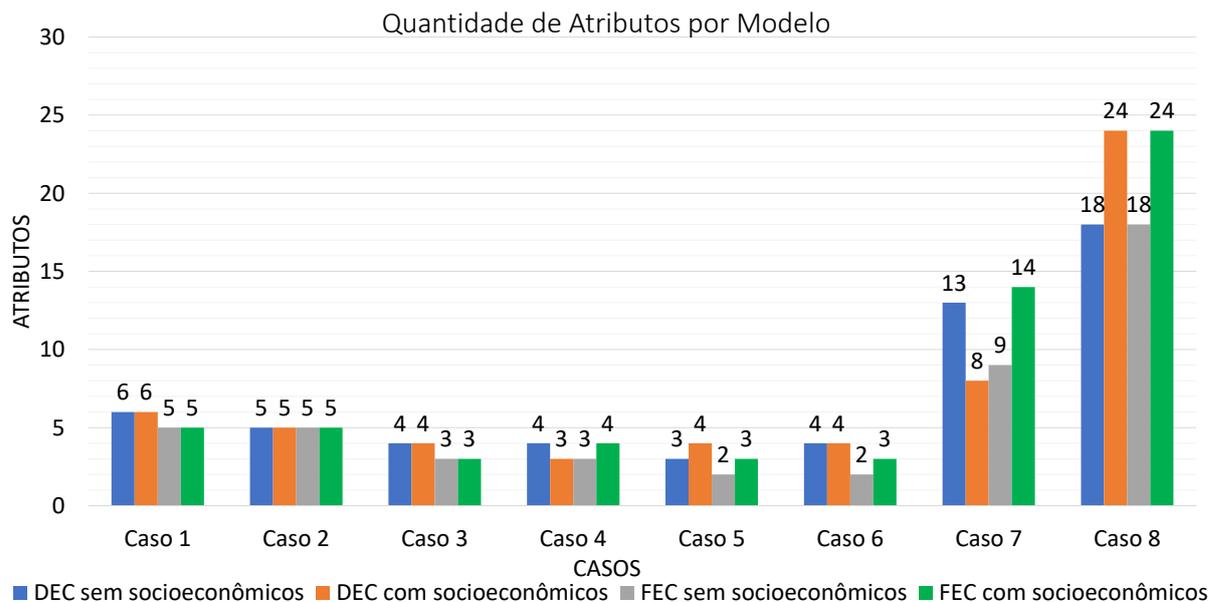
Tabela 4.20 – Parâmetros utilizados nos estudos de caso.

Estudo de Caso	Correlação mínima	VIF máximo	Técnica de regressão
1	0,2	5	<i>Stepwise</i>
2	0,2	3	<i>Stepwise</i>
3	0,3	5	<i>Stepwise</i>
4	0,3	3	<i>Stepwise</i>
5	0,5	5	<i>Stepwise</i>
6	0,5	3	<i>Stepwise</i>
7	0,2	5	LASSO
8	0,2	5	RIDGE

Fonte: do Autor.

A Figura 4.2 apresenta graficamente a quantidade de atributos presentes nos modelos definidos através dos estudos de caso.

Figura 4.2 – Quantidade de Atributos por Modelo



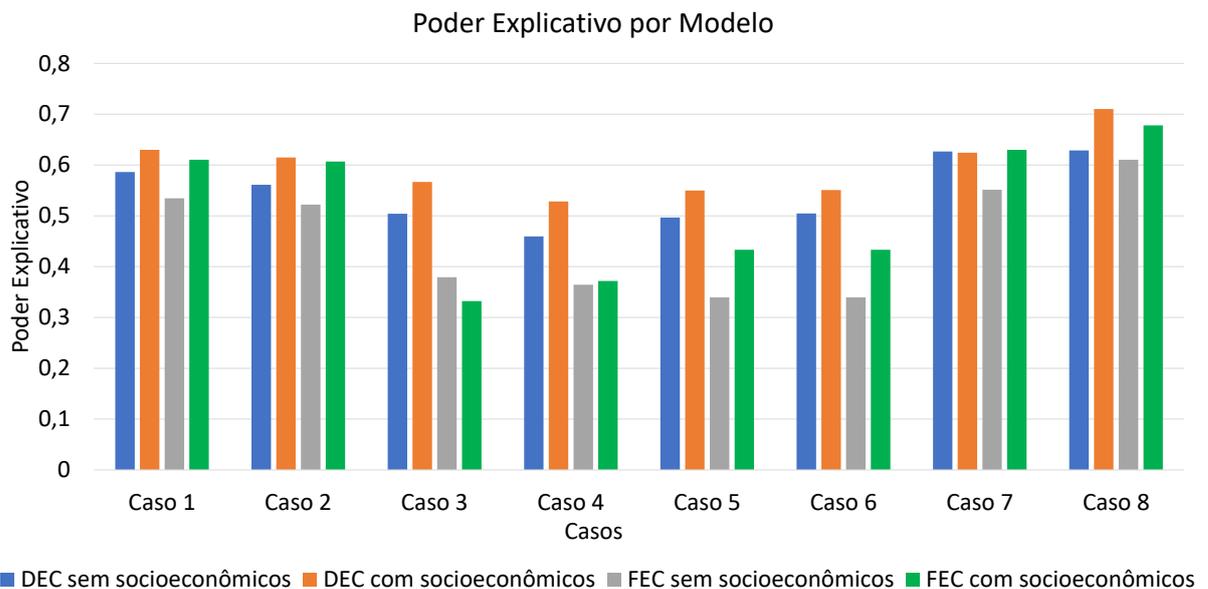
Fonte: do Autor

Analisando a Figura 4.2, verifica-se que nos casos 1, 3 e 5, onde foi utilizada a Regressão *Stepwise* com diferentes limites de correlação, diminuiu o número de atributos

presentes nos modelos na medida que o limite de correlação aumentou. Nos casos onde houve a redução do VIF máximo (estudos de caso 2, 4 e 6), observa-se que não ocorreram alterações expressivas na quantidade de variáveis. Além disso, o caso 8, onde foi empregada a Regressão *Ridge*, possui a maior quantidade de atributos, uma vez que a técnica de regressão utilizada não realiza a exclusão de variáveis do modelo.

O poder explicativo de cada modelo gerado nos estudos de caso é apresentado na Figura 4.3.

Figura 4.3 – Poder Explicativo por Modelo

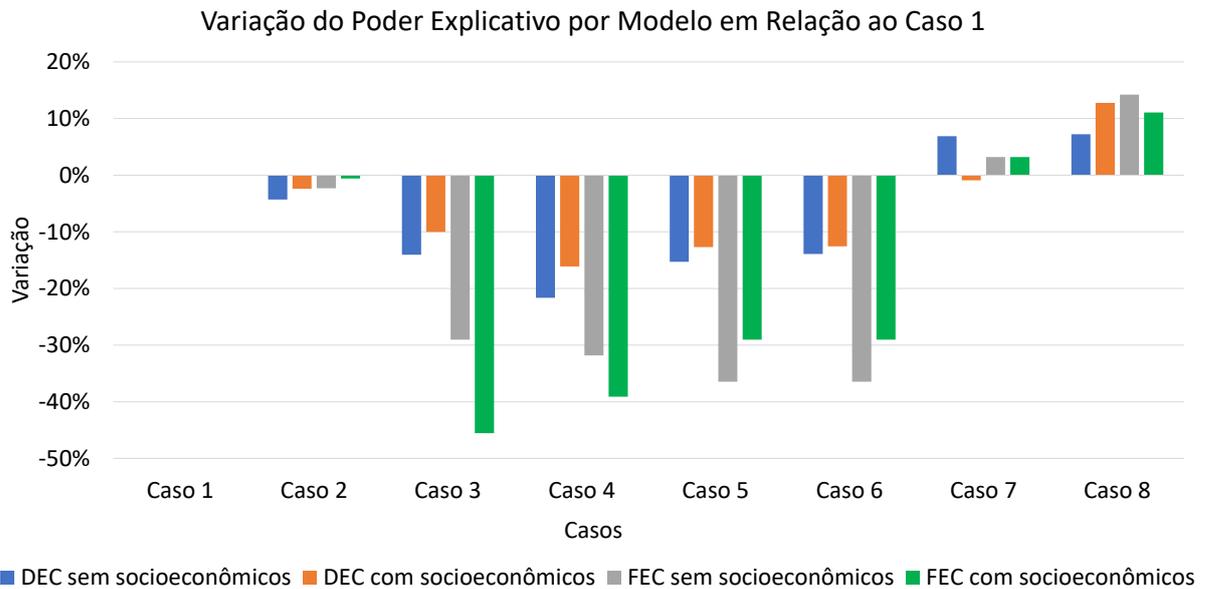


Fonte: do Autor

Observando a Figura 4.3, percebe-se que além de possuir a maior quantidade de variáveis, o estudo de caso 8 também possui os modelos com o maior poder explicativo. Em relação aos casos onde foi utilizada a regressão *Stepwise*, nota-se que o aumento do limite mínimo de correlação causou uma redução no poder explicativo dos modelos, enquanto que a diminuição do limite máximo do VIF não causou alterações significativas no poder explicativo dos modelos.

É importante investigar a variação do poder explicativo dos modelos em relação ao estudo de caso 1, uma vez que este estudo de caso utiliza os mesmos parâmetros aplicados pela ANEEL para definir o modelo vigente. Desta forma, a Figura 4.4 apresenta a variação do poder explicativo por modelo em relação ao caso 1.

Figura 4.4 – Variação do Poder Explicativo por Modelo em Relação ao Caso 1

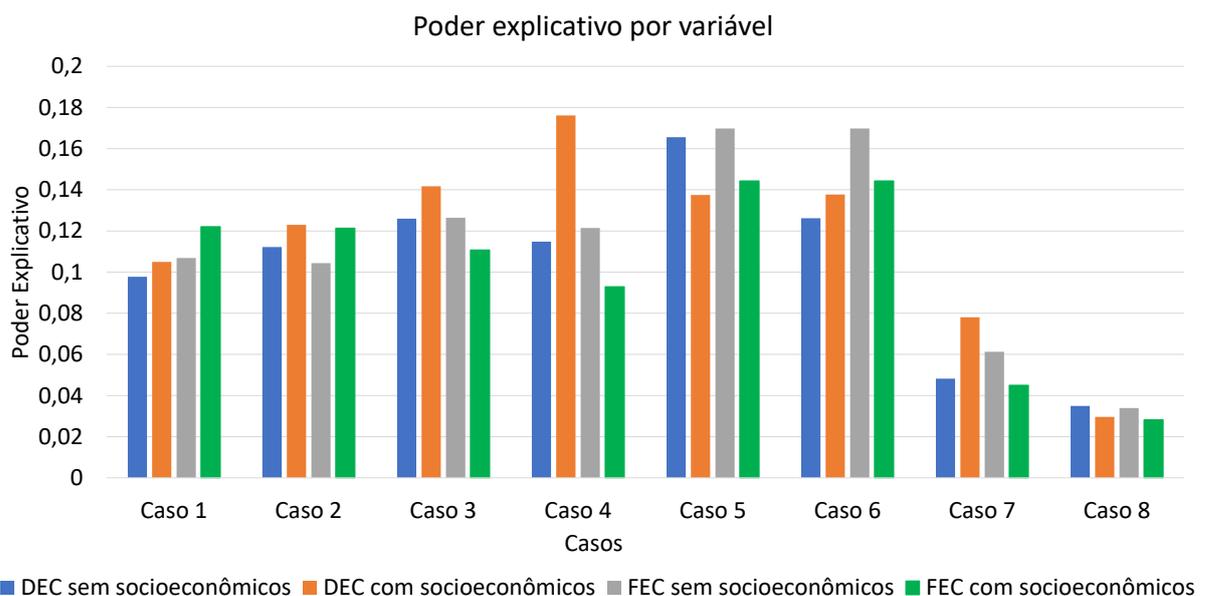


Fonte: do Autor

Analisando a Figura 4.4, é possível notar que os estudos de caso que apresentaram as maiores variações negativas são os que utilizam a regressão *Stepwise* com diferentes critérios de VIF e de correlação, sendo que os estudos de caso 3 e 4, onde é utilizado a correlação limite de 30%, apresentaram as variações mais significativas.

A Figura 4.5 expõe o poder explicativo por variável presente nos modelos.

Figura 4.5 – Poder Explicativo por Variável

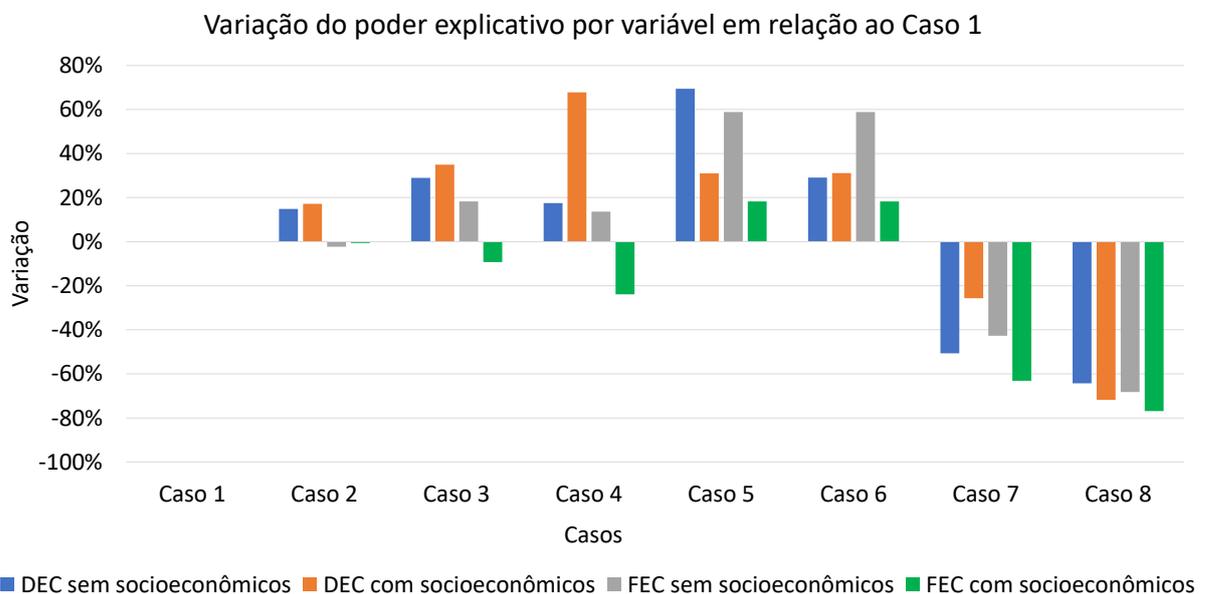


Fonte: do Autor

Observando a Figura 4.5, pode-se averiguar que ao aumentar o limite de correlação nos modelos que utilizam a Regressão *Stepwise*, apesar de haver uma redução no poder explicativo do modelo, ocorreu um aumento no poder explicativo por variável. Isto denota que o poder explicativo parcial dos atributos destes modelos é maior que dos modelos do estudo de caso 1. Além disso, percebe-se que nos estudos de caso 7 e 8, onde foram utilizadas as regressões LASSO e Ridge, respectivamente, o poder explicativo por variável é pequeno, uma vez que estes modelos possuem uma quantidade elevada de atributos.

Além do mais, é relevante verificar a variação do poder explicativo por variável dos modelos gerados em relação ao modelo do estudo de caso 1. Assim sendo, a Figura 4.6 apresenta a variação do poder explicativo por variável em relação ao caso 1.

Figura 4.6 – Variação do Poder Explicativo por Variável em Relação ao Caso 1



Fonte: do Autor

Analisando a variação do poder explicativo por variável tomando como referência o estudo de caso 1, constata-se que a maior variação positiva, considerando os modelos sem variáveis socioeconômicas, está no estudo de caso 5, onde foi utilizado um limite mínimo de correlação igual a 40% e um limite máximo do VIF igual a 5.

4.10 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os estudos de caso e os resultados obtidos ao aplicar a análise estatística para seleção dos atributos que caracterizam os indicadores coletivos. As simulações foram realizadas no *software SAS University Edition*. Foram elaborados 8 estudos de caso variando-se os limites de correlação e VIF e a técnica de

regressão.

Através dos dados obtidos, verificou-se que, quando se utiliza a regressão *Stepwise*, ao aumentar o limite de correlação ocorre uma redução no poder explicativo do modelo e na quantidade de variáveis, no entanto, há um aumento no poder explicativo por variável. Desta forma, aumentar o limite de correlação gera modelos mais parcimoniosos, isto é, que explicam a variável resposta com menos variáveis dependentes.

Em relação a utilização das regressões LASSO e Ridge, nota-se que houve um aumento expressivo na quantidade de variáveis no modelo enquanto que o poder explicativo aumentou de forma não significativa.

5 CONCLUSÃO

Neste trabalho foi apresentada uma análise estatística dos atributos que caracterizam os indicadores de continuidade coletivos. Esta análise está estruturada da mesma forma que a análise estatística realizada pela ANEEL, isto é, foram seguidas as mesmas etapas com mudanças nos critérios como correlação, VIF e a técnica de regressão. Desta forma, esta análise proposta auxilia na investigação de atributos que possam estar mais relacionados com as causas de interrupções do fornecimento de energia elétrica.

A validação da análise proposta deu-se por meio de oito estudos de caso, onde foram analisados os efeitos da mudança dos critérios anteriormente citados. Em cada estudo de caso, foram gerados dois modelos para cada indicador de continuidade, em que um modelo foi gerado considerando as variáveis socioeconômicas e o outro desconsiderando, assim como na análise conduzida pela ANEEL. Desta forma, foram concebidos 4 modelos por estudo de caso.

O primeiro estudo de caso consistiu em replicar os critérios adotados pela ANEEL. Dos quatro modelos construídos neste caso, apenas o modelo do indicador DEC sem variáveis socioeconômicas foi congruente com o modelo da ANEEL, os demais diferenciaram-se pelas variáveis presentes no modelo e/ou pela quantidade de variáveis. Este fato pode ser explicado pela baixa replicabilidade que a regressão *Stepwise* apresenta, uma vez que foram utilizados os mesmos dados da análise da ANEEL.

Utilizando a regressão *Stepwise* foram elaborados seis estudos de caso, onde os parâmetros modificados foram o critério de correlação e o VIF. Verificou-se que aumentar o limite máximo de correlação causou uma redução no número de variáveis e no poder explicativo dos modelos, no entanto, aumentou o poder explicativo por atributo presente no modelo. Portanto, de um ponto de vista estatístico, aumentar o limite de correlação, diminui o poder explicativo total, mas seleciona variáveis mais relacionadas com os indicadores.

Quanto a redução do limite do VIF, não houve mudanças significativas no poder explicativo ou na quantidade de variáveis presentes no modelo. Os modelos gerados ao utilizar as regressões LASSO e Ridge foram os que apresentaram o maior poder explicativo e a maior quantidade de variáveis, portanto, ao analisar o poder explicativo por variável, pode-se concluir que estas regressões geraram modelos com variáveis pouco relacionadas com os indicadores.

5.1 TRABALHOS FUTUROS

Existem diferentes formas de se definir os modelos dos indicadores de continuidade, seja aplicando outras técnicas ou variando-se apenas os critérios de análise. Este

trabalho apresenta uma contribuição ao processo contínuo de procura por atributos que sejam relacionados com os indicadores de continuidade. Portanto, sugere-se como trabalho futuros:

- Utilização de algoritmos de *Machine Learning* para construção dos modelos;
- Aplicação da análise estatística com atributos distintos dos fornecidos pela ANEEL;
- Implementação no *software* MATLAB de uma análise estatística integrada com a *clusterização* dos conjuntos.

5.2 TRABALHOS PUBLICADOS

Durante a elaboração deste trabalho de conclusão de curso, foram desenvolvidos os seguintes trabalhos:

5.2.1 Artigos publicados em anais de eventos

1. **SANTOS, J. R;** SILVEIRA, L.; SANTOS, L. L. C.; TISCHER, C. B.; GARCIA, R. A. F. Método de Extração de Atributos Técnicos e Geográficos de Distribuidoras de Energia Elétrica In: 13th Seminar on Power Electronics and Control (SEPOC 2021), 2021, Santa Maria - Brasil.

2. SILVEIRA, L.; **SANTOS, J. R;** SANTOS, L. L. C.; GARCIA, R. A. F. Análise de Sensibilidade do Método Dinâmico na Definição de Limites dos Indicadores de Continuidade In: Conferência Brasileira sobre Qualidade da Energia Elétrica (XIV CBQEE), 2021, Foz do Iguaçu - Brasil.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, M. F.; LOTUFO, A. D. P.; LOPES, M. L. M. Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, v. 1, n. 1, 2013.

ANEEL, A. N. D. E. E. **Resolução Normativa nº24**. Brasília: [s.n.], 2000. Acesso em 31 jan. 2021. Disponível em: <<http://www2.aneel.gov.br/cedoc/res2000024.pdf>>.

_____. **Nota Técnica n.028/2010**. Brasília: [s.n.], 2010. Acesso em 29 jan. 2021. Disponível em: <https://www2.aneel.gov.br/aplicacoes/consulta_publica/documentos/Nota%20T%C3%A9cnica%2028_2010%20-%20SRD%20-%20Anexo.pdf>.

_____. **Nota Técnica n.102/2014**. Brasília: [s.n.], 2014. Acesso em 14 jan. 2021. Disponível em: <https://www2.aneel.gov.br/aplicacoes/audiencia/arquivo/2014/029/resultado/nota_tecnica_0102_2014_srd.pdf>.

_____. **Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional PRODIST Módulo 8 Qualidade da Energia Elétrica**. Brasília: [s.n.], 2021. Acesso em 29 jan. 2021. Disponível em: <https://www.aneel.gov.br/documents/656827/14866914/M%C3%B3dulo_8-Revis%C3%A3o_12/342ff02a-8eab-2480-a135-e31ed2d7db47>.

FAN, L. et al. Variable selection and model prediction based on lasso, adaptive lasso and elastic net. In: **2015 4th International Conference on Computer Science and Network Technology (ICCSNT)**. [S.l.: s.n.], 2015. v. 01, p. 579–583.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5. ed. Porto Alegre: AMGH Editora, 2011.

HASTIE, T.; TIBSHIRANI, R.; WAINWRIGHT, M. **Statistical learning with sparsity: the lasso and generalizations**. [S.l.]: CRC press, 2015.

HEUMANN, C.; SCHOMAKER, M. et al. **Introduction to statistics and data analysis**. [S.l.]: Springer, 2016.

HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, Taylor & Francis Group, v. 12, n. 1, p. 55–67, 1970.

KAGAN, N.; ROBBA, E. J.; SCHMIDT, H. P. **Estimação de indicadores de qualidade da energia elétrica**. [S.l.]: Editora Blucher, 2009.

Lin, C.; Lin, C. Using ridge regression method to reduce estimation uncertainty in chlorophyll models based on worldview multispectral data. In: **IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium**. [S.l.: s.n.], 2019. p. 1777–1780.

Luo, H.; Liu, Y. A prediction method based on improved ridge regression. In: **2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)**. [S.l.: s.n.], 2017. p. 596–599.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. 5. ed. Nova Jersey: John Wiley & Sons, 2012.

MUTHUKRISHNAN, R.; ROHINI, R. Lasso: A feature selection technique in predictive modeling for machine learning. In: **2016 IEEE International Conference on Advances in Computer Applications (ICACA)**. [S.l.: s.n.], 2016. p. 18–20.

PESSANHA, J. F. M. **Um Modelo de Análise Envoltória de Dados para Estabelecimento das Metas de Continuidade do Fornecimento de Energia Elétrica**. 2006. 161 f. Tese (Doutorado em Engenharia Elétrica) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006.

RAWLINGS, J. O.; PANTULA, S. G.; DICKEY, D. A. **Applied Regression Analysis: A Research Tool**. 2. ed. Nova York: Springer, 2011.

SAS. **SAS university Edition**. 2021. Acesso em 25 jul. 2021. Disponível em: <https://www.sas.com/pt_br/software/university-edition.html>.

SCHIELD, M. Correlation, determination and causality in introductory statistics. **ASA Proceedings of the Section on Statistical Education**, v. 1995, p. 189–194, 08 1995.

SMITH, G. Step away from stepwise. **Journal of Big Data**, SpringerOpen, v. 5, n. 1, p. 1–12, 2018.

TANURE, J. E. P. S. **Análise comparativa de empresas de distribuição para o estabelecimento de metas de desempenho para indicadores de continuidade do serviço de distribuição**. 2000. 157 f. Dissertação (Mestrado em Engenharia Elétrica) — Escola Federal de Engenharia de Itajubá, Itajubá, 2000.

TANURE, J. E. P. S. **Proposta de procedimentos e metodologia para estabelecimento de Metas de Qualidade (DEC e FEC) para concessionárias de distribuição de energia elétrica através da análise comparativa**. 2004. 228 f. Tese (Doutorado em Engenharia Elétrica) — Universidade de São Paulo, São Paulo, 2004.

THOMPSON, B. **Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial**. [S.l.]: Sage Publications Sage CA: Thousand Oaks, CA, 1995.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 58, n. 1, p. 267–288, 1996.

ANEXO A – 146 ATRIBUTOS

Tabela A.1 – 146 Atributos

NOME DO ATRIBUTO (unidade)	SIGLA DO ATRIBUTO
ÁREA (km ²)	AREA
ÁREA DE ATUAÇÃO (km ²)	AA
ARRUAMENTO (km)	ARR
DENSIDADE DE ARRUAMENTO (km/km ²)	DENS_ARR
DENSIDADE DE ARRUAMENTO POR ÁREA DE ATUAÇÃO (km/km ²)	DENS_ARR_AA
ÁREA DE ATUAÇÃO ATENDIDA POR ARRUAMENTO - BUFFER (km ²)	AA_ATEND
ÁREA DE ATUAÇÃO NÃO ATENDIDA POR ARRUAMENTO - BUFFER (km ²)	AA_NATEND
PERCENTUAL DA ÁREA DE ATUAÇÃO ATENDIDA POR RUAS E ESTRADAS - BUFFER (%)	PC_AA_ATEND
RODOVIAS NÃO PAVIMENTADAS (km)	EST_NPAV
RODOVIAS PAVIMENTADAS (km)	EST_PAV
RODOVIAS (km)	EST
DENSIDADE DE RODOVIAS NÃO PAVIMENTADAS (km/km ²)	DENS_EST_NPAV
DENSIDADE DE RODOVIAS PAVIMENTADAS (km/km ²)	DENS_EST_PAV
DENSIDADE DE RODOVIAS (km/km ²)	DENS_EST
DENSIDADE DE RODOVIAS NÃO PAVIMENTADAS POR ÁREA DE ATUAÇÃO (km/km ²)	DENS_EST_NPAV_AA
DENSIDADE DE RODOVIAS PAVIMENTADAS POR ÁREA DE ATUAÇÃO (km/km ²)	DENS_EST_PAV_AA
DENSIDADE DE RODOVIAS POR ÁREA DE ATUAÇÃO (km/km ²)	DENS_EST_AA
PERCENTUAL DE RODOVIAS NÃO PAVIMENTADAS (%)	PC_EST_NPAV
PERCENTUAL DE RODOVIAS NO ARRUAMENTO TOTAL (%)	PC_EST_ARR
ÁREA DO BUFFER DE ARRUAMENTO NÃO PAVIMENTADO (km ²)	A_ARR_NPAV
ÁREA DO BUFFER DE ARRUAMENTO PAVIMENTADO (km ²)	A_ARR_PAV
PERCENTUAL DE ÁREA DO BUFFER DE ARRUAMENTO NÃO PAVIMENTADO NA ÁREA TOTAL (%)	PC_A_ARR_NPAV

Continua na próxima página

Tabela A.1 – *Continuação*

NOME DO ATRIBUTO (unidade)	SIGLA DO ATRIBUTO
PERCENTUAL DE ÁREA DO BUFFER DE ARRUAMENTO PAVIMENTADO NA ÁREA TOTAL (%)	PC_A_ARR_PAV
DENSIDADE DE DESCARGAS ATMOSFÉRICAS (raios/km ²)	D_DESC
PRECIPITAÇÃO PLUVIOMÉTRICA MÉDIA ANUAL (mm)	PLUV
PRECIPITAÇÃO PLUVIOMÉTRICA MÉDIA ANUAL NA ÁREA DE ATUAÇÃO (mm)	PLUV_AA
ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE ALTA (km ²)	VRA_AA
PERCENTUAL DE ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE ALTA (%)	PC_VRA_AA
ÁREA COM VEGETAÇÃO REMANESCENTE ALTA (km ²)	VRA
PERCENTUAL DE ÁREA COM VEGETAÇÃO REMANESCENTE ALTA (%)	PC_VRA
ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE BAIXA (km ²)	VRB_AA
PERCENTUAL DE ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE BAIXA (%)	PC_VRB_AA
ÁREA COM VEGETAÇÃO REMANESCENTE BAIXA (km ²)	VRB
PERCENTUAL DE ÁREA COM VEGETAÇÃO REMANESCENTE BAIXA (%)	PC_VRB
ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE MÉDIA (km ²)	VRM_AA
PERCENTUAL DE ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE MÉDIA (%)	PC_VRM_AA
ÁREA COM VEGETAÇÃO REMANESCENTE MÉDIA (km ²)	VRM
PERCENTUAL DE ÁREA COM VEGETAÇÃO REMANESCENTE MÉDIA (%)	PC_VRM
ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE ALTA OU MÉDIA (km ²)	VRAM_AA
PERCENTUAL DE ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE ALTA OU MÉDIA (%)	PC_VRAM_AA
ÁREA COM VEGETAÇÃO REMANESCENTE ALTA OU MÉDIA (km ²)	VRAM
PERCENTUAL DE ÁREA COM VEGETAÇÃO REMANESCENTE ALTA OU MÉDIA (%)	PC_VRAM

Continua na próxima página

Tabela A.1 – *Continuação*

NOME DO ATRIBUTO (unidade)	SIGLA DO ATRIBUTO
ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE (km ²)	VR_AA
PERCENTUAL DE ÁREA DE ATUAÇÃO COM VEGETAÇÃO REMANESCENTE (%)	PC_VR_AA
ÁREA COM VEGETAÇÃO REMANESCENTE (km ²)	VR
PERCENTUAL DE ÁREA COM VEGETAÇÃO REMANESCENTE (%)	PC_VR
NUC POR LOCALIZAÇÃO URBANA	NUC_URB
NUC POR LOCALIZAÇÃO NÃO URBANA	NUC_NURB
NUC TOTAL	NUC
PERCENTUAL DE NUC URBANO (%)	PC_NUC_URB
DENSIDADE DE UNIDADES CONSUMIDORAS (UC/km ²)	NUC_AREA
DENSIDADE DE UNIDADES CONSUMIDORAS POR ÁREA DE ATUAÇÃO (UC/km ²)	NUC_AA
NUC POR CLASSE RESIDENCIAL	NUC_RES
NUC POR CLASSE COMERCIAL	NUC_COM
NUC POR CLASSE INDUSTRIAL	NUC_IND
NUC POR CLASSE RURAL	NUC_RUR
NUC POR CLASSE PODER PÚBLICO	NUC_PP
NUC POR CLASSE ILUMINAÇÃO PÚBLICA	NUC_IP
NUC POR CLASSE CONSUMO PRÓPRIO	NUC_CP
SERVIÇO PÚBLICO	NUC_SP
NUC POR CLASSE COMERCIAL E OUTRAS CLASSES	NUC_COM_OU
NUC POR OUTRAS CLASSES	NUC_OU
PERCENTUAL DE NUC RESIDENCIAL (%)	PC_NUC_RE
PERCENTUAL DE NUC COMERCIAL (%)	PC_NUC_COM)
PERCENTUAL DE NUC INDUSTRIAL (%)	PC_NUC_IND
PERCENTUAL DE NUC RURAL (%)	PC_NUC_RUR
PERCENTUAL DE NUC COMERCIAL E OUTRAS CLASSES (%)	PC_NUC_COM_OU
PERCENTUAL DE NUC OUTRAS CLASSES (%)	PC_NUC_OU
ENERGIA ANUAL POR CLASSE RESIDENCIAL - INDQUAL (MWh)	ENE_RES

Continua na próxima página

Tabela A.1 – *Continuação*

NOME DO ATRIBUTO (unidade)	SIGLA DO ATRIBUTO
ENERGIA ANUAL POR CLASSE COMERCIAL - INDQUAL (MWh)	ENE_COM
ENERGIA ANUAL POR CLASSE INDUSTRIAL - INDQUAL (MWh)	ENE_IND
ENERGIA ANUAL POR CLASSE RURAL- INDQUAL (MWh)	ENE_RUR
ENERGIA ANUAL POR OUTRAS CLASSES - INDQUAL (MWh)	ENE_OU
ENERGIA ANUAL POR CLASSE COMERCIAL E OUTRAS CLASSES - INDQUAL (MWh)	ENE_COM_OU
ENERGIA TOTAL ANUAL (MWh)	ENE
CONSUMO MÉDIO ANUAL POR UC (MWh)	CM_NUC
DENSIDADE DE CONSUMO (MWh/km ²)	ENE_AREA
DENSIDADE DE CONSUMO POR ÁREA DE ATUAÇÃO (MWh/km ²)	ENE_AA
CONSUMO MÉDIO POR UC POR CLASSE RESIDENCIAL (MWh)	CM_NUC_RES)
CONSUMO MÉDIO POR UC POR CLASSE COMERCIAL (MWh)	CM_NUC_COM
CONSUMO MÉDIO POR UC POR CLASSE INDUSTRIAL (MWh)	CM_NUC_IND
CONSUMO MÉDIO POR UC POR CLASSE RURAL (MWh)	CM_NUC_RUR
CONSUMO MÉDIO POR UC POR OUTRAS CLASSES (MWh)	CM_NUC_OU
CONSUMO MÉDIO POR UC POR CLASSE COMERCIAL E OUTRAS CLASSES (MWh)	CM_NUC_COM_OU
PERCENTUAL DE ENERGIA DA CLASSE RESIDENCIAL (%)	PC_ENE_RES
PERCENTUAL DE ENERGIA DA CLASSE COMERCIAL (%)	PC_ENE_COM
PERCENTUAL DE ENERGIA DA CLASSE INDUSTRIAL (%)	PC_ENE_IND
PERCENTUAL DE ENERGIA DA CLASSE RURAL (%)	PC_ENE_RUR
PERCENTUAL DE ENERGIA DE OUTRAS CLASSES (%)	PC_ENE_OU
PERCENTUAL DE ENERGIA DA CLASSE COMERCIAL E OUTRAS CLASSES (%)	PC_ENE_COM_OU
REDES MT MONOFÁSICAS (km)	ERMT_1F
REDES MT BIFÁSICAS (km)	ERMT_2F
REDES MT TRIFÁSICAS (km)	ERMT_3F

Continua na próxima página

Tabela A.1 – *Continuação*

NOME DO ATRIBUTO (unidade)	SIGLA DO ATRIBUTO
PERCENTUAL DE REDES MT MONOFÁSICAS (%)	PC_ERMT_1F
PERCENTUAL DE REDES MT TRIFÁSICAS (%)	PC_ERMT_3F
REDES MT POR LOCALIZAÇÃO URBANA (km)	ERMT_URB
REDES MT POR LOCALIZAÇÃO NÃO URBANA (km)	ERMT_NURB
REDES MT (km)	ERMT
PERCENTUAL DE REDES MT NÃO URBANAS (%)	PC_ERMT_NURB
DENSIDADE DE REDES MT (km/km ²)	ERMT_AREA
DENSIDADE DE REDES MT POR ÁREA DE ATUAÇÃO (km/km ²)	ERMT_AA
QUANTIDADE DE ALIMENTADORES MT	ALIM
COMPRIMENTO MÉDIO MT POR ALIMENTADOR (km)	COMP_M_ALIM
DENSIDADE DE UC POR REDE MT (UC/km)	NUC_ERMT
DENSIDADE DE ENERGIA POR REDE MT (MWh/km)	ENE_ERMT
TRANSFORMADORES DE DISTRIBUIÇÃO MONOFÁSICOS	TD_1F
TRANSFORMADORES DE DISTRIBUIÇÃO BIFÁSICOS	TD_2F
TRANSFORMADORES DE DISTRIBUIÇÃO TRIFÁSICOS	TD_3F
PERCENTUAL DE TRANSFORMADORES MONOFÁSICOS (%)	PC_TD_1F
PERCENTUAL DE TRANSFORMADORES TRIFÁSICOS (%)	PC_TD_3F)
POTÊNCIA NOMINAL INSTALADA MONOFÁSICA (kVA)	PNI_1F
POTÊNCIA NOMINAL INSTALADA BIFÁSICA (kVA)	PNI_2F
POTÊNCIA NOMINAL INSTALADA TRIFÁSICA (kVA)	PNI_3F
PERCENTUAL DA POTÊNCIA INSTALADA EM TRANSFORMADORES MONOFÁSICOS (%)	PC_PNI_1F
PERCENTUAL DA POTÊNCIA INSTALADA EM TRANSFORMADORES TRIFÁSICOS (%)	PC_PNI_3F
TRANSFORMADORES DIST POR LOCALIZAÇÃO URBANO	TD_URB
TRANSFORMADORES DIST POR LOCALIZAÇÃO NÃO URBANA	TD_NURB
PERCENTUAL DE TRANSFORMADORES NÃO URBANOS (%)	PC_TD_NURB
TRANSFORMADORES DIST	TD
DENSIDADE DE TRANSFORMADORES POR ÁREA (TD/km ²)	TD_AREA

Continua na próxima página

Tabela A.1 – *Continuação*

NOME DO ATRIBUTO (unidade)	SIGLA DO ATRIBUTO
DENSIDADE DE TRANSFORMADORES POR ÁREA DE ATUAÇÃO (TD/km ²)	TD_AA
UC POR TRANSFORMADOR	NUC_TD
POTÊNCIA NOMINAL MÉDIA POR TRANSFORMADOR (kVA)	PNI_TD
DENSIDADE DE TRANSFORMADORES POR km DE REDE MT (TD/km)	TD_ERMT
POTÊNCIA NOMINAL INSTALADA POR LOCALIZAÇÃO URBANA (kVA)	PNI_URB
POTÊNCIA NOMINAL INSTALADA POR LOCALIZAÇÃO NÃO URBANA (kVA)	PNI_NURB
PERCENTUAL DA POTÊNCIA INSTALADA EM TRANSFORMADORES NÃO URBANOS (%)	PC_PNI_NURB
POTÊNCIA NOMINAL INSTALADA TOTAL (kVA)	PNI
DENSIDADE DE POTÊNCIA NOMINAL INSTALADA (kVA/km ²)	PNI_AREA
DENSIDADE DE POTÊNCIA NOMINAL INSTALADA POR ÁREA DE ATUAÇÃO (kVA/km ²)	PNI_AA
DENSIDADE DE POTÊNCIA NOMINAL INSTALADA POR km DE REDE MT (kVA/km)	PNI_ERMT
POTÊNCIA NOMINAL MÉDIA POR UC (kVA)	PNI_NUC
ENERGIA CONSUMIDA POR POTÊNCIA NOMINAL INSTALADA (MWh/kVA)	ENE_PNI
Índice de Gini	GINI
% de extremamente pobres	PEXP
Renda per capita	RDPC
% da população em domicílios com banheiro e água encanada	BAN_ENC
% de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados	AGUA_ESG_IN
% de pessoas em domicílios com paredes inadequadas	PAREDE_IN
Índice de Desenvolvimento Humano Municipal	IDHM
Cobertura de água - total	AGUA
Coleta de lixo - total	LIXO
3 SM - % de chefes de família com renda de até 3 salários mínimos	SM3

Continua na próxima página

Tabela A.1 – *Continuação*

NOME DO ATRIBUTO (unidade)	SIGLA DO ATRIBUTO
Precariedade - % de pessoas em domicílios subnormais (Censo 2010)	SUB
Violência - Óbitos por agressão	VIO
Violência - Óbitos por agressão por 100 mil habitantes	VIO_100

Fonte: (ANEEL, 2014)