

## Análise Estatística para Seleção dos Atributos que Caracterizam os Indicadores de Continuidade do Fornecimento de Energia Elétrica

Jaderson Rosa dos Santos\* Leonardo da Silveira\*\*  
Laura Lisiane Callai dos Santos\*\*\* Rui Anderson Ferrarezi Garcia\*\*\*\*

\* Universidade Federal de Santa Maria, RS,  
(e-mail: r.jaderson@gmail.com).

\*\* Universidade Federal de Santa Maria, RS,  
(e-mail: leonardo.dasilveira@outlook.com)

\*\*\* Universidade Federal de Santa Maria, RS,  
(e-mail: laura.santos@ufsm.br)

\*\*\*\* Companhia Estadual de Distribuição de Energia Elétrica, RS,  
(e-mail: ruifg@ceee.equatorialenergia.com.br)

---

**Abstract:** To assess the quality of service provided by distributors, the National Electric Energy Agency (ANEEL) establishes service continuity indicators and limits for these indicators. The methodology for defining the limits is constantly updated, aiming to make it consistent with the characteristics of the distributors. In this context, the present work aims to analyze the change in the criteria of the statistical analysis for the selection of 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 the characterization of continuity indicators, however, different statistical techniques from those employed by ANEEL are used. To validate the proposed methodology, eight case studies were developed using the *software SAS University Edition*. It was found that, using *Stepwise* regression, the increase in the correlation limit 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.

**Resumo:** Para avaliar a qualidade do serviço prestado pelas distribuidoras, a Agência Nacional de Energia Elétrica (ANEEL) estabelece indicadores de continuidade do serviço e limites para estes indicadores. A metodologia de definição dos limites é constantemente atualizada, visando torná-la coerente com as características das distribuidoras. 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*. 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.

*Keywords:* Quality; Indicators; Attributes; Limits; Continuity.

*Palavras-chaves:* Qualidade; Indicadores; Atributos; Limites; Continuidade.

---

### 1. INTRODUÇÃO

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 do

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. 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 nº 46/1978 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 nº 024/2000, 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 nº 102/2014 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 influenciam 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 aos indicadores de continuidade coletivos.

## 2. REVISÃO BIBLIOGRÁFICA

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.

### 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) 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).

Existem quatro indicadores individuais que são apurados por unidade consumidora. São definidos dois indicadores coletivos que são apurados a partir dos indicadores individuais.

A metodologia de análise comparativa utilizada pela ANEEL baseia-se na técnica de regulação por desempenho *Yardstick Competition*, proposta por 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*.

Para definir a semelhança dos conjuntos, elaborou-se um modelo com seis variáveis para cada indicador coletivo, sendo que cinco variáveis são idênticas para ambos os modelos, como mostra a Tabela 1. Estas variáveis compreendem características físico-elétricas dos conjuntos de unidades consumidoras.

Tabela 1. Variáveis Seleccionadas para o Modelo da ANEEL.

DEC	FEC
Percentual de Número de Unidades consumidoras por localização Urbana (%)	Percentual de Número de Unidades consumidoras por localização Urbana (%)
Percentual de Área com Vegetação Remanescente Alta ou Média (%)	Percentual de Área com Vegetação Remanescente Alta ou Média (%)
Precipitação Pluviométrica Média Anual (mm)	Precipitação Pluviométrica Média Anual (mm)
Percentual de Redes Média Tensão Trifásicas (%)	Percentual de Redes Média Tensão Trifásicas (%)
Número de Unidades Consumidoras da Classe Industrial	Número de Unidades Consumidoras da Classe Comercial

Os modelos foram gerados a partir da eliminação de variáveis utilizando coeficientes de correlação seguida da regressão de 146 atributos oriundos da Base de Dados Geográficos (BDGD) das distribuidoras. A regressão utilizada foi a *Stepwise*.

### 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).

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).

### 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 et al., 2012). 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 and Porter, 2011).

A Regressão *Ridge* foi introduzida por (Hoerl and Kennard, 1970) como um aperfeiçoamento ao Método dos Mínimos Quadrados Ordinários (MMQO). Neste método, utiliza-se uma normalização L2 para limitar a minimização do resíduo, que é uma variável aleatória que exprime a diferença entre os valores da reta gerada pela regressão e os valores observados da variável resposta, 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 em (1).

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

Em que:

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

A limitação do resíduo, também chamada de penalidade, é controlada pela variável  $\lambda$ .

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

A regressão LASSO, assim como a regressão *ridge*, é uma otimização do MMQO, entretanto, uma regularização L1 é aplicada 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 et al. (2015). A equação do método LASSO é expressa em (2).

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

Em que:

- $\hat{\beta}_{lasso}$ : coeficiente de regressão LASSO;

- $\lambda$ : 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 and Rohini (2016).

A Regressão *Stepwise* é uma combinação da Inclusão Passo a Frente e da Eliminação Passo Atrás, 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 et al. (2012). Em outras palavras, um regressor pode perder significância no modelo na medida em que outros regressores são adicionados ao modelo.

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 Soma dos Quadrados dos Resíduos (SQR) e a variação do modelo em relação à média das observações da variável dependente, também chamada de soma total de quadrados (STQ) Gujarati and Porter (2011). Logo, a representação matemática do  $R^2$  é dada por (3).

$$R^2 = \frac{SQR}{STQ} = \frac{\sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

Em que:

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

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 et al. (2011). A forma matemática para se obter o VIF é dada por (4).

$$VIF_j = \frac{1}{1 - R_j^2} \quad (4)$$

Em que:

VIF<sub>j</sub>: VIF da variável j;

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

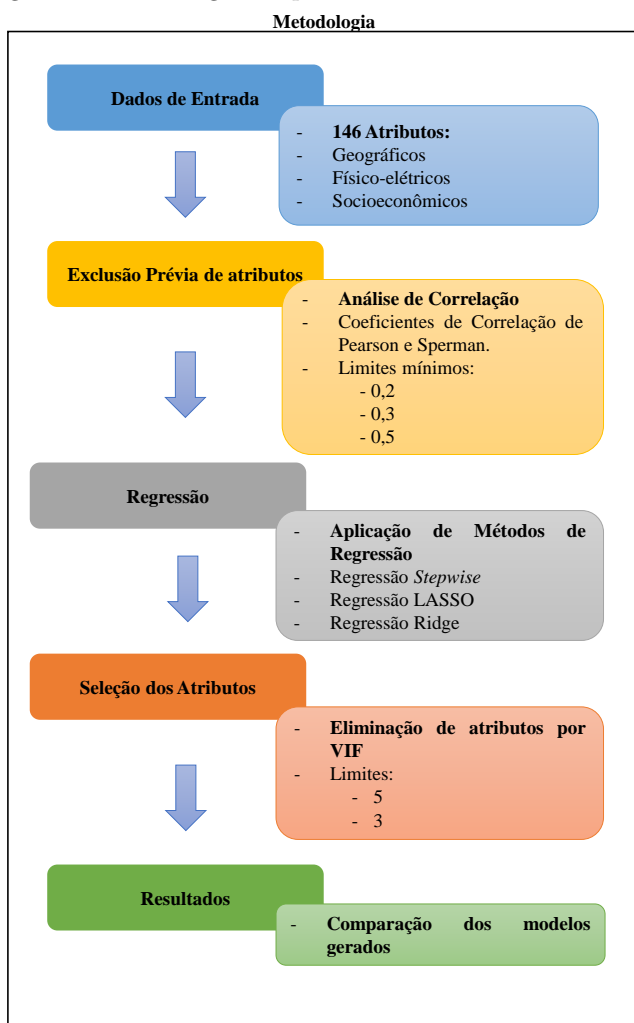
Essas técnicas estatísticas apresentadas foram empregadas na metodologia proposta neste trabalho.

### 3. METODOLOGIA

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 1.

Figura 1. Metodologia Proposta.



Conforme Figura 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.

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.

Tendo em posse os atributos candidatos ao modelo após a filtragem prévia, a próxima etapa, de acordo com a Figura 1, consiste em realizar a regressão destas variáveis. As técnicas de regressão utilizadas para construir os modelos são: Regressão *Stepwise*, Regressão LASSO e Regressão Ridge. 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.

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, que permite mensurar a capacidade de um modelo teórico representar a variável dependente.

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. Ao analisar o poder explicativo por modelo, é possível inferir a capacidade do conjunto de variáveis presentes no modelo de prever a variável dependente, enquanto que o poder explicativo por variável indica a relevância individual de cada atributo presente no modelo.

Para a definição dos modelos teóricos foi utilizado o *software SAS University Edition*, desenvolvido pela empresa de inteligência analítica SAS, para construir os estudos de caso.

### 4. ESTUDO DE CASO

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.

#### 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 conforme Figura 1, visando eliminar os atributos que possuem correlação de Pearson ou Spearman com DEC ou FEC inferior a 0,2, conforme mostra a Tabela 2.

Tabela 2. 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>

Para o modelo do DEC considerando as variáveis socioeconômicas e desconsiderando as variáveis socioeconômicas, remanesceram 69 e 58 variáveis independentes, respectivamente. Estas variáveis foram submetidas a seleção de atributos visando eliminar a multicolinearidade. 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*, conforme a terceira etapa da Metodologia mostrada na Figura 1, resultando em 17 variáveis independentes para o DEC com atributos socioeconômicos e 9 atributos para o DEC sem atributos socioeconômicos. A Tabela 3 apresenta os resultados da regressão *Stepwise* para os modelos do DEC considerando variáveis socioeconômicas.

Tabela 3. Resultado da regressão *Stepwise* para o modelo do DEC considerando variáveis socioeconômicas.

Com Variáveis Socioeconômicas			
N	Atributo	R <sup>2</sup>	R <sup>2</sup> Modelo
1	PAREDE_IN	0,3255	0,3255
2	PC_NUC_URB	0,1634	0,4889
3	PC_ERMT_3F	0,0492	0,5381
4	PLUV	0,0432	0,5813
5	PC_NUC_COM	0,0264	0,6077
6	PC_VRAM	0,0222	0,6298
7	CM_NUC_RES	0,0115	0,6414
8	PC_NUC_RUR	0,0094	0,6507
9	IDHM	0,0089	0,6596
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	PNLNURB	0,0009	0,6945

Na Tabela 4 são apresentados os resultados da regressão *Stepwise* para os modelos do DEC considerando variáveis socioeconômicas.

Tabela 4. Resultado da regressão *Stepwise* para o modelo do DEC desconsiderando variáveis socioeconômicas.

Sem Variáveis Socioeconômicas			
N	Atributo	R <sup>2</sup>	R <sup>2</sup> Modelo
1	PC_NUC_URB	0,3249	0,3249
2	PC_VRAM	0,0905	0,4153
3	PC_ERMT_3F	0,0590	0,4743
4	PLUV	0,0462	0,5205
5	CM_NUC_RES	0,0405	0,5610
6	NUC_IND	0,0252	0,5862
7	PC_NUC_COM	0,0122	0,5985
8	NUC_RUR	0,0080	0,6065
9	ERMT_AREA	0,0097	0,6162

Observando as Tabela 3 e 4, 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 5.

Tabela 5. Modelos Selecionados para o DEC.

Com Varáveis Socioeconômicas		Sem Varáveis Socioeconômicas	
Atributos	R <sup>2</sup>	Atributos	R <sup>2</sup>
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	

Nota-se que as variáveis selecionadas para o modelo do DEC sem atributos socioeconômicos são idênticas às variáveis utilizados no modelo da ANEEL, apresentado na Tabela 1. Assim como foi definido os modelos do DEC, deve-se definir os modelos do 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. Assim como ocorreu na análise do DEC, percebe-se que os modelos com variáveis socioeconômicas possuem mais variáveis que modelos sem variáveis socioeconômicas.

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, conforme mostra a Tabela 6.

Tabela 6. Resultado da regressão *Stepwise* para o modelo do FEC considerando variáveis socioeconômicas.

Com Variáveis Socioeconômicas			
N	Atributo	R <sup>2</sup>	R <sup>2</sup> Modelo
1	LIXO	0,3563	0,3563
2	PLUV	0,0957	0,4520
3	ERMT_AREA	0,0825	0,5345
4	PEXP	0,0408	0,5753
5	NUC_COM	0,0350	0,6103
6	ERMT_1F	0,0184	0,6287
7	PC_VRAM	0,0202	0,6489
8	TD_1F	0,0204	0,6694
9	PAREDE_IN	0,0079	0,6773
10	IDHM	0,0048	0,6821
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

Foram atribuídos 10 atributos para o modelo do FEC sem atributos socioeconômicos, conforme mostra a Tabela 7.

Tabela 7. Resultado da regressão *Stepwise* para o modelo do FEC desconsiderando variáveis socioeconômicas.

Sem Variáveis Socioeconômicas			
N	Atributo	R <sup>2</sup>	R <sup>2</sup> Modelo
1	TD_ERMT	0,2218	0,2218
2	PLUV	0,1484	0,3702
3	PC_VRAM	0,0984	0,4686
4	PC_NUC_URB	0,0451	0,5137
5	CM_NUC	0,0207	0,5343
6	PC_NUC_RES	0,0137	0,5480
7	PC_ERMT_3F	0,0113	0,5593
8	NUC_RUR	0,0039	0,5632
9	CM_NUC_RES	0,0016	0,5648
10	DENS_EST_PAV	0,0018	0,5666

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 8.

Tabela 8. Modelos Selecionados para o FEC.

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

Comparando o modelo sem variáveis socioeconômicas selecionado com o modelo da ANEEL, apresentado na Tabela 1, nota-se que há uma variável a menos e, das 5 variáveis selecionadas, 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

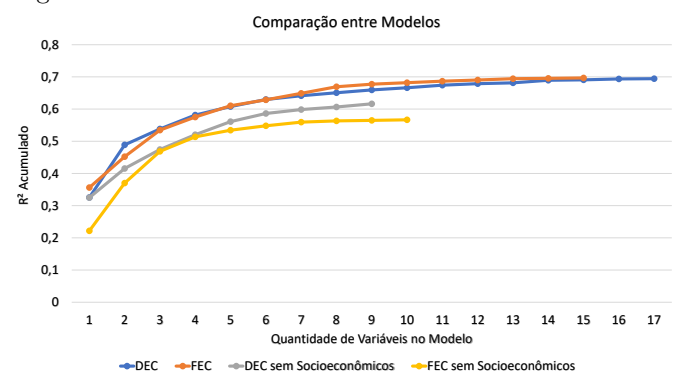
e esta, por sua vez, é realizada por com base nos atributos que fazem parte do modelo do indicador.

#### 4.2 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 2, 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 2. Modelos do Caso 1



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

#### 4.3 Comparação dos Modelos

Os modelos gerados neste trabalho seguiram os mesmos procedimentos do Estudo de Caso 1 para manter as mesmas etapas do processo de seleção de variáveis adotado pela ANEEL.

Após a construção de todos os modelos, é necessários compará-los, visando inferir como o poder explicativo e a quantidade de variáveis é modificada pela alteração de determinados parâmetros estatísticos.

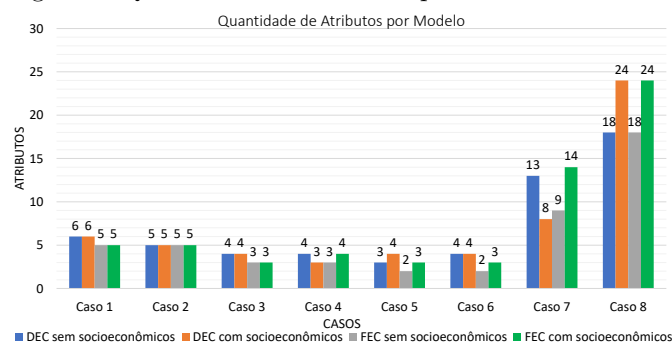
Os parâmetros utilizados em todos os estudos de caso são apresentados na Tabela 9, destacando em cinza os que são idênticos a análise da ANEEL.

Tabela 9. Parâmetros utilizados nos estudos de caso.

Estudo de Caso	Correlação mínima	VIF máximo	Regressão
1	0,2	5	Stepwise
2	0,2	3	Stepwise
3	0,3	5	Stepwise
4	0,3	3	Stepwise
5	0,5	5	Stepwise
6	0,5	3	Stepwise
7	0,2	5	LASSO
8	0,2	5	RIDGE

A Figura 3 apresenta graficamente a quantidade de atributos presentes nos modelos definidos através dos estudos de caso.

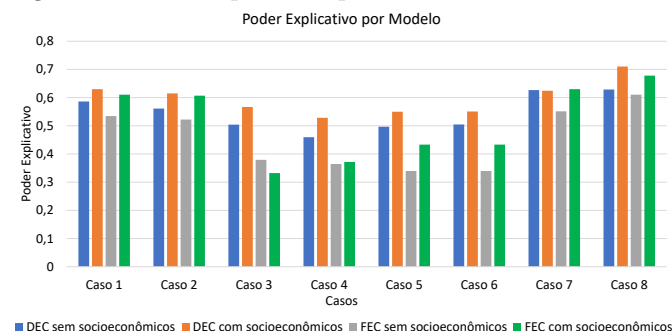
Figura 3. Quantidade de Atributos por Modelo



Analisando a Figura 3, verifica-se que nos casos 1, 3 e 5, onde foi utilizada a Regressão *Stepwise*, 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 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.

Figura 4. Poder Explicativo por Modelo

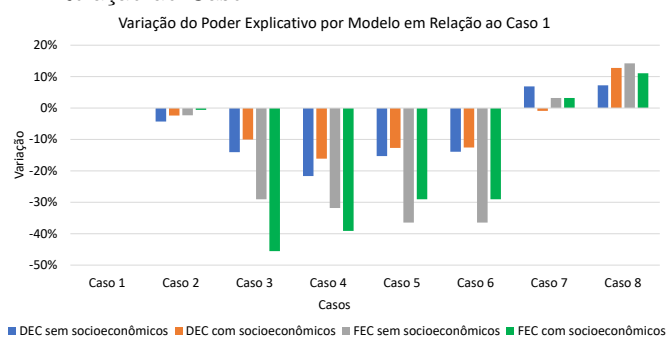


Observando a Figura 4, percebe-se que além de possuir a maior quantidade de variáveis, o estudo de caso 8 também possui os modelos com o maior poder explicativo. Em relação aos casos onde foi utilizada a regressão *Stepwise*, nota-se que o aumento do limite mínimo de correlação causou uma redução no 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 5 apresenta a variação do poder explicativo por modelo em relação ao caso 1.

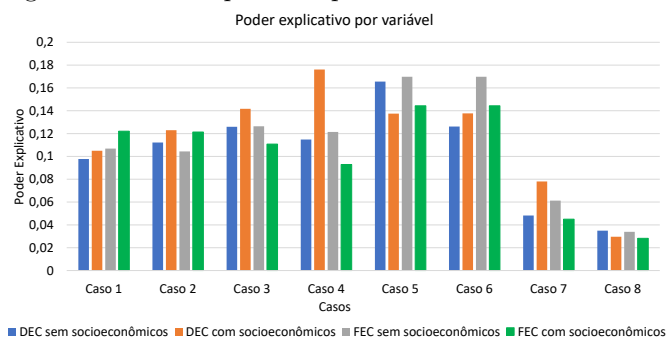
Figura 5. Variação do Poder Explicativo por Modelo em Relação ao Caso 1



Analisando a Figura 5, 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 6 expõe o poder explicativo por variável presente nos modelos.

Figura 6. Poder Explicativo por Variável



Observando a Figura 6, 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 7 apresenta a variação do poder explicativo por variável em relação ao caso 1.



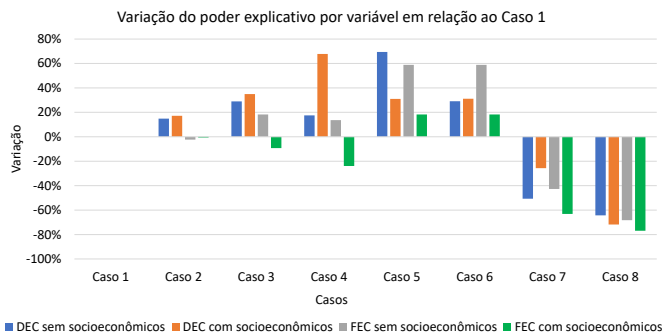
## 6. AGRADECIMENTOS

Ao Programa de P&D regulado pela ANEEL e executado pela Companhia Estadual de Distribuição de Energia Elétrica (CEEE-D) via chamada n°01/2018.

## REFERÊNCIAS

- ANEEL, A.N.D.E.E. (2000). Resolução normativa n°24. URL <http://www2.aneel.gov.br/cedoc/res2000024.pdf>. Acesso em 31 jan. 2021.
- ANEEL, A.N.D.E.E. (2014). Nota técnica n.102/2014. URL [https://www2.aneel.gov.br/aplicacoes/audiencia/arquivo/2014/029/resultado/nota\\_tecnica\\_0102\\_2014\\_srd.pdf](https://www2.aneel.gov.br/aplicacoes/audiencia/arquivo/2014/029/resultado/nota_tecnica_0102_2014_srd.pdf). Acesso em 14 jan. 2021.
- ANEEL, A.N.D.E.E. (2021). Procedimentos de distribuição de energia elétrica no sistema elétrico nacional – prodist módulo 8 – qualidade da energia elétrica. URL [https://www.aneel.gov.br/documents/656827/14866914/M\C3%B3dulo\\_8-Revis\C3%A3o\\_12/342ff02a-8eab-2480-a135-e31ed2d7db47](https://www.aneel.gov.br/documents/656827/14866914/M\C3%B3dulo_8-Revis\C3%A3o_12/342ff02a-8eab-2480-a135-e31ed2d7db47). Acesso em 29 jan. 2021.
- Gujarati, D.N. and Porter, D.C. (2011). *Econometria Básica*. AMGH Editora, Porto Alegre, 5 edition.
- Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.
- Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.
- Montgomery, D.C., Peck, E.A., and Vining, G.G. (2012). *Introduction to linear regression analysis*. John Wiley & Sons, Nova Jersey, 5 edition.
- Muthukrishnan, R. and Rohini, R. (2016). Lasso: A feature selection technique in predictive modeling for machine learning. In *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, 18–20.
- Pessanha, J.F.M. (2006). *Um Modelo de Análise Envoltória de Dados para Estabelecimento das Metas de Continuidade do Fornecimento de Energia Elétrica*. Doutorado em engenharia elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- Rawlings, J.O., Pantula, S.G., and Dickey, D.A. (2011). *Applied Regression Analysis: A Research Tool*. Springer, Nova York, 2 edition.
- Schild, M. (1995). Correlation, determination and causality in introductory statistics. *ASA Proceedings of the Section on Statistical Education*, 1995, 189–194.
- Tanure, J.E.P.S. (2000). *Análise comparativa de empresas de distribuição para o estabelecimento de metas de desempenho para indicadores de continuidade do serviço de distribuição*. Mestrado em engenharia elétrica, Escola Federal de Engenharia de Itajubá, Itajubá.
- Tanure, J.E.P.S. (2004). *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*. Doutorado em engenharia elétrica, Universidade de São Paulo, São Paulo.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

Figura 7. Variação do Poder Explicativo por Variável em Relação ao Caso 1



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.

## 5. CONCLUSÃO

Neste trabalho foi apresentada uma análise estatística dos atributos que caracterizam os indicadores de continuidade coletivos. A validação da análise proposta deu-se por meio de oito estudos de caso, onde foram analisados os efeitos da mudança dos critérios estatísticos. Em cada estudo de caso, foram gerados dois modelos para cada indicador de continuidade, considerando e desconsiderando as variáveis socioeconômicas, assim como na análise conduzida pela ANEEL.

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.

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.