Resumo expandido

Em 2 de agosto de 2022 a Agência Nacional de Energia Elétrica (ANEEL) publicou no Diário Oficial da União (DOU), seção 3, página 106, a abertura da Tomada de Subsídios – TS nº 14/2022. A TS 014/2022 tem com o objetivo obter subsídios e informações acerca de base de dados que será utilizada no estudo de benchmarking dos custos operacionais regulatórios dos agentes de transmissão. Sucintamente, para o cálculo dos custos operacionais regulatórios é aplicado um modelo conhecido como Análise Envoltória de Dados (Data Envelopment Analysis) utilizando o custo operacional como insumo e nove produtos. Os produtos representam os principais direcionadores (drivers) para os custos operacionais das empresas brasileiras de transmissão de energia elétrica. Nesse sentido, é possível assumir que exista uma forte correlação entre os produtos e o insumo, podendo ser aplicado um modelo estatístico de regressão múltipla para estimar parâmetros e realizar inferências estatísticas sobre o modelo. Utilizando a abordagem estatística, o insumo (custo operacional) representa a variável resposta/dependente (y) e os produtos representam as variáveis preditoras (x\_1,x\_2,…,x\_p). Entretanto, a presença de forte correlação linear entre os produtos gera valores inconsistentes para os coeficientes do modelo de regressão como a estimativa de um coeficiente negativo quando é esperado, do ponto de vista técnico, um coeficiente positivo. A literatura apresenta possíveis soluções para esse problema como parametrizações alternativas para o modelo de regressão ou o uso de modelos Bayesianos. Uma solução alternativa é o uso de programação linear para a estimação dos parâmetros do modelo de regressão impondo restrições na forma de inequações. Este trabalho utiliza modelo de regressão linear, programação linear e técnicas de validação cruzada para avaliar os principais direcionadores dos custos operacionais das empresas brasileiras de transmissão de energia elétrica. Os resultados indicam que, com a aplicação da metodologia de programação linear e do uso de restrições aos parâmetros do modelo, apenas 2 dentre os 9 produtos são identificados como os principais direcionadores dos custos operacionais. Dessa forma, este resultado indica que a atual metodologia para o cálculo dos custos operacionais eficientes utiliza direcionadores redundantes, podendo comprometer ou enviesar estimativas de eficiência.

**Resumo**

A Agência Nacional de Energia Elétrica (ANEEL) publicou a Tomada de Subsídios – TS nº 14/2022 para a obtenção de informações sobre a base de dados que será utilizada no estudo de benchmarking dos custos operacionais regulatórios dos agentes de transmissão. O cálculo desses custos é feito por meio da Análise Envoltória de Dados (Data Envelopment Analysis), utilizando o custo operacional como insumo e nove produtos como variáveis explicativas (drivers) dos custos. No entanto, devido à forte correlação linear entre as variáveis explicativas, o modelo de regressão pode gerar valores inconsistentes para os coeficientes estimados. Uma alternativa para lidar com esse problema é a utilização de programação linear e restrições nas inequações para estimação dos parâmetros do modelo. Nesse contexto, este trabalho propõe uma metodologia que combina a regressão linear com a programação linear e técnicas de validação cruzada para avaliar os principais direcionadores dos custos operacionais das empresas brasileiras de transmissão de energia elétrica. Os resultados indicam que apenas duas das nove variáveis explicativas são identificadas como os principais direcionadores dos custos operacionais, o que sugere a presença de variáveis redundantes na metodologia atual de cálculo dos custos operacionais eficientes, podendo comprometer as estimativas de eficiência.

Palavras-chave: Regressão linear. Programação linear. Bootstrap não-paramétrico. Custo operacional. Transmissão de energia elétrica.

**Abstract**

The Brazilian National Electricity Agency (ANEEL) published the Public Consultation - PC No. 14/2022 to obtain information about the database to be used in the benchmarking study of regulatory operating costs for transmission agents. The calculation of these costs is carried out using Data Envelopment Analysis (DEA), with operating costs as input and nine products as explanatory variables (drivers) of costs. However, due to the strong linear correlation between the explanatory variables, the regression model may generate inconsistent values for the estimated coefficients. An alternative to address this issue is to use linear programming and inequality constraints to estimate the model parameters. In this context, this work proposes a methodology that combines linear regression with linear programming and cross-validation techniques to evaluate the main drivers of operating costs for Brazilian electricity transmission companies. The results indicate that only two out of the nine explanatory variables are identified as the main drivers of operating costs, suggesting the presence of redundant variables in the current methodology for calculating efficient operating costs, which may compromise efficiency estimates.

Key words: Linear regression. Linear programming. Non-parametric bootstrap. Operational cost. Transmission of electricity.

**1 Introdução**

O sistema energético brasileiro (SEB) é composto por empresas públicas e privadas que são reguladas pela Agência Nacional de Energia Elétrica (ANEEL). O Art. 2º da Lei nº 9.427 de 26 de dezembro de 1996 (Brasil, 1996) dispõe que a ANEEL tem como finalidade regular e fiscalizar a produção, transmissão, distribuição e comercialização de energia elétrica. É também de responsabilidade a ANEEL estabelecer e rever as tarifas segundo do Art. 3º, Incisos XII e XXII (Brasil, 1996).

Para os agentes envolvidos no SEB, a revisão de tarifas é de importância central para a determinação das receitas anuais permitidas (RAP) das instalações de transmissão. Segundo Pessanha et al (2010), a RAP envolve a apuração dos custos anuais dos ativos elétricos, dos custos de administração e custos operacionais, acrescidos dos tributos aplicáveis do setor e de uma parcela de ajuste. Costa et al (2022) discorre que tais custos operacionais compreendem uma pequena parte da tarifa de energia e do custo operacional regulatório, dessa forma é crucial que tal parte seja calculada de forma apropriada, para que o operador do sistema possa cobrar o consumidor final de forma adequada.

Miranda Lopes et al (2016) argumenta que uma das dificuldades do setor elétrico no Brasil envolve fornecer aos consumidores finais energia com preços justos com remuneração justa para o distribuidor. Tal remuneração é fortemente afetada pelos custos operacionais do setor, por isso o regulador estimula que os partícipes do sistema promovam redução deles.

Para Pessanha et al (2010), o desafio do cálculo dos custos operacionais eficientes do setor é a assimetria de informação entre o regulador e os agentes do mercado. Para evitar tal assimetria e definir tais custos de forma procedimental, a reguladora por meio da Resolução Normativa - REN ANEEL n° 257 de 6 de março de 2007 (Brasil, 2007) estabeleceu os conceitos gerais e procedimentos para revisão da primeira revisão tarifária periódica (RTP). A partir de então, em cada ciclo de revisão a reguladora divulga documentos normativos e técnicos com orientações para os procedimentos a serem realizados.

Em fevereiro de 2017, a reguladora emitiu a Superintendência de Regulação Econômica e Estudos do Mercado (SRM) emitiu a Nota Técnica nº 037/2017-SEM/ANEEL por meio da qual apresentou as bases de dados que seriam utilizadas no estudo de eficiência dos custos operacionais das concessionárias de transmissão e que subsidiaria a definição dos custos operacionais regulatórios dos contratos de concessão que passariam por revisão no período de julho de 2018 a junho de 2023.

A ANEEL publicou a Nota Técnica nº 160/2017 (SRM) (Brasil,2017) que dispõe sobre as regras de apuração dos custos operacionais regulatórios. A reguladora explicita no documento publicado que vem adotando a prática de realizar análises comparativas entre os custos operacionais dos operadores do SEB, tornando possível a atribuição de um nível de eficiência a cada empresa e, então, definir os valores de custos operacionais regulatórios associados a uma referência de eficiência. Na nota, fica definido que o insumo utilizado no estudo foi a despesa operacional das transmissoras, composto pelas contas de pessoal, materiais, serviços de terceiros, seguros, tributos e outros (PMSO).

No ano de 2018 foi publicado pela ANEEL Nota Técnica nº 204/2018-SRM/ANEEL (Brasil, 2018) que apresenta a metodologia a ser utilizada no cálculo dos custos operacionais regulatórios no âmbito do processo de revisão periódica das receitas anuais permitidas das concessionárias de transmissão de energia elétrica. Conforme descrito no documento, é um ponto central de todo o processo da definição dos custos operacionais o estabelecimento de critério para aferição do nível eficiência de custos das empresas, por isso a reguladora vinha discutindo com a sociedade o modelo de benchmarking a ser utilizado visto que há a necessidade de se levar em consideração as características e particularidades de cada agente. Posteriormente serão discutidos a metodologia, variáveis e demais ajustes definidos no documento.

A tomada de subsídio é uma etapa crucial para o ciclo regulatório para a revisão periódica da RAP. Dessa forma a reguladora publicou nota a Nota Técnica nº 97/2018-SRM/ANEEL para consolidação das bases de dados e informações que serão utilizadas para embasar o benchmarking dos custos operacionais regulatórios das concessionárias de transmissão. Visando dar transparência ao processo, foi publicado no Diário Oficial da União (DOU), seção 3, página 106, a abertura da Tomada de Subsídios – TS nº 14/2022.

Em prosseguimento a TS nº 14/2022, a reguladora publicou em 29 de julho de 2022 a Nota Técnica Nº 97/2022–SRM/ANEEL com objetivo de propor instauração de Tomada de Subsídios para apresentar e consolidar as bases de dados que serão utilizadas para subsidiar o estudo de benchmarking de eficiência dos custos operacionais regulatórios das concessionárias de transmissão que passarão por revisão periódica da Receita Anual Permitida – RAP em 1º de julho de 2023, conforme data contratual. A NT 97/2022 buscou dar publicidade e transparência para a base de dados que será utilizada, permitindo que a sociedade a avalie, critique, audite e valide, viabilizando a realização de eventuais correções que se mostrem necessárias e conferindo maior transparência ao processo como um todo. A base de dados foi posteriormente publicada conforme descrição das notas técnicas.

**1.1 Justificativa**

Segundo Feitosa Neto (2009), a determinação correta da RAP para as concessionárias atuantes no SEB é de extrema importância para não comprometer a operação do sistema e não onerar excessivamente o consumidor final. Dessa forma, apurar de forma robusta e realista o custo operacional (PMSO) é crucial para o cálculo da RAP e para manutenção do nível tarifário.

A implementação de uma técnica estatística que visa corrigir distorções matemáticas e qualitativas e que modele as restrições operacionais dos direcionadores de custo é de extrema importância para que se evite o comprometimento das estimativas de eficiência dos custos operacionais. Este trabalho busca contribuir para as discussões acerca do tema e para oferecer ao regulador uma alternativa para o cálculo do parâmetro tão considerável para as definições regulatórias do SEB.

**1.2 Objetivos**

O presente trabalho tem como propósito implementar um modelo de previsão que respeite as restrições técnicas impostas pela natureza operacional dos direcionadores de custo e avaliar seus resultados, implicações e compará-lo com modelos tradicionais. Portanto, deseja-se que variáveis preditoras (direcionadores de custo) do modelo contribuam de forma positiva ou neutra com a variável resposta (PMSO), eliminando qualquer incongruência ou deturpação da realidade imposta.

São objetivos específicos nessa pesquisa:

* Analisar as correlações existentes entre as variáveis e também sua representatividade e importância para a realização de previsões.
* Definir um modelo de regressão linear múltipla ou de programação linear para o custo operacional que respeite as restrições impostas pela natureza da operação.
* Discutir os resultados encontrados e apresentar uma alternativa ao regulador para o modelo utilizado na definição dos custos operacionais.

**2 Referencial Teórico**

**2.1 Análise de regressão**

Segundo Upton et al (2014), é o modelo de regressão estatístico mais simples e mais utilizado. Montgomery et al (2003) definem a análise de regressão como uma técnica estatística utilizada para investigar a relação de uma ou mais variáveis. Para Hoffman e Vieira (1977) a análise de regressão permite conhecer os efeitos quer umas variáveis exercem ou que pareçam exercer sobre outras, os autores ainda classificam que tal método é o mais importante para a econometria.

O método de regressão linear foi formalizado por Francis Galton em 1891 em sua publicação Hereditary Genius*,* onde o autor realiza um estudo da relação entre as alturas de pais e filhos. De forma sucinta, o autor concluiu com os resultados observados por sua técnica que pais altos tendem a ter filhos altos e pais baixos tendem a ter filhos mais baixos. Segundo Stanton (2001), após as contribuições de Karl Pearson em 1896, 1922 e 1930 o método obteve rigor matemático e foi amplamente utilizado.

**2.1.1 Regressão Linear Univariada**

Montgomery et al (2003) define que a regressão linear univariada ou regressão linear simples consiste na relação entre uma variável aleatória e outra de acordo com a expressão abaixo, onde a é a variável dependente, é a variável aleatória independente, é o coeficiente da regressão chamado de intercepto e é o coeficiente da regressão chamado de coeficiente de inclinação.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Em termos formais, Montgomery et al (2003) definem que a relação entre e seja uma linha reta onde o valor esperado para cada observação em cada nível de seja uma variável aleatória. Em outros termos, os autores explicam que a maneira apropriada de se generalizar tal afirmação para um modelo linear probabilístico é considerar que o valor esperado de seja em função de , mas que para um valor fixo de o valor real de seja determinado pela função do valor médio mais um termo de erro aleatório, onde representa tal termo com média zero e variância :

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

**2.1.2 Método de Mínimos Quadrados e propriedades dos estimadores**

Para implementação do modelo linear probabilístico proposto anteriormente, é preciso realizar a estimação do coeficiente angular ou de inclinação () e o intercepto (). O método dos mínimos quadrados (MQO) é um dos métodos mais utilizados para estimação os coeficientes de regressão linear univariada. Em termos gerais o método consiste na minimização da soma dos quadrados das diferenças entre a variável dependente observada no conjunto de dados de entrada e a saída da função da variável independente, entretanto para que o MQO entregue estimadores consistentes, Lewis-Back (2015) e Kennedy (2002) defendem que determinados pressupostos devem ser satisfeitos:

1. A esperança do termo de erro deve ser igual a zero () e tem distribuição normal com variância .
2. Deve haver homoscedasticidade dos termos de erro, ou seja, a variância do erro é constante para os diferentes valores da variável independente
3. As variáveis independentes não podem apresentar alta correlação entre si, ou seja, não podem apresentar a propriedade de multicolinearidade
4. A variável independente não pode apresentar correlação com o termo de erro

Segundo Montgomery et al (2003), para estimação dos coeficientes, a soma dos quadrados dos desvios das observações em relação a linha de regressão dada por:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Dessa forma os estimadores de mínimos quadrados devem satisfazer:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Após simplificação:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Finalmente, as estimativas dos coeficientes podem ser descritas como:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

A equação da reta ajustada para o MQO:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

**2.1.3 Regressão Linear Múltipla**

Se na regressão linear univariada tínhamos apenas uma única variável independente para a modelagem, nesse caso o modelo poderá ter termos que se relacionam com a variável dependente. Miloni et al (1995) dissertam que tal método é indicado para realizar investigações em casos em que várias variáveis afetam simultaneamente a variável de interesse, podendo ser uma ferramenta poderosa.

Para representar a o modelo de regressão linear múltipla, podemos utilizar a seguinte equação, onde representa a quantidade de variáveis independentes, representa o intercepto da equação e o efeito da variável independente.

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

Para estimação dos coeficientes da regressão no modelo multivariado é possível utilizar o MQO ajustado conforme descreve Montgomery et al (2003):

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

Minimizando com relação aos coeficientes, as estimativas têm que satisfazer:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

As equações originadas a partir de (10) são chamadas de equações normais de mínimos quadrados:

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

Pode-se reescrever o modelo de regressão múltipla em forma de notação matricial, no qual há variáveis independentes, observações. No caso, representa o vetor (x1) de observações da variável dependente, representa uma matriz ( x ), o vetor ( x ) com os coeficientes do modelo

|  |  |  |
| --- | --- | --- |
|  |  | (12) |

Sendo:

|  |  |  |
| --- | --- | --- |
|  |  | (13) |

**2.1.4 Propriedades, pressupostos e adequação dos modelos de regressão linear**

Para validar o ajuste de um modelo de regressão linear é preciso verificar sua representatividade e respeitar os pressupostos de normalidade dos resíduos, assumindo que devem apresentar variância constante e média igual a zero.

Sharpe et al (2011) apresentam algumas suposições e condições para o modelo de regressão linear múltipla. Primeiramente é a suposição de linearidade (significância do modelo), é importante verificar se há uma relação linear entre as variáveis explicativas e a variável resposta. Isso pode ser feito plotando o diagrama de dispersão de y versus cada variável X e verificando se há uma relação linear. Também é recomendado plotar o diagrama dos resíduos para detectar violações das condições de linearidade. Tal suposição pode ser verificada pelo teste de significância.

A suposição de independência dos erros do modelo de regressão deve ser verificada para garantir que eles são independentes entre si. Isso significa que os erros não devem estar correlacionados. Além disso, é importante que os dados sejam provenientes de uma amostra aleatória ou experimento aleatório para garantir a condição de aleatoriedade.

Outro ponto importante é suposição de igualdade das variâncias dos erros. Como falado anteriormente, tal suposição deve ser verificada para garantir que a variabilidade dos erros é aproximadamente a mesma para todos os valores das variáveis independentes. Por fim, a suposição de normalidade dos erros é importante para garantir que sigam uma distribuição normal

Em relação ao erro aleatório , é preciso realizar a estimação de sua variância para realização dos testes de significância do modelo. Tais testes são válidos para as regressões univariada e multivariada pois tem o mesmo propósito. Montgomery et al (2003) explicam que os testes de significância são os balizadores para definição se o modelo estimado é adequado ou não, tanto

Para estimar o termo da variância do erro, pode-se realizar o cálculo da soma dos quadrados dos resíduos . A consiste na soma das diferenças ao quadrado do termo das observações reais () com os valores estimados .

|  |  |  |
| --- | --- | --- |
|  |  | (14) |

Assim como os resíduos, é possível calcular a soma dos quadrados da regressão que consiste na soma das diferenças ao quadrado dos valores estimados da média dos valores observados ().

|  |  |  |
| --- | --- | --- |
|  |  | (15) |

De posse dos termos calculados, podemos obter a soma dos quadrados totais da forma abaixo e reescrever quaisquer temos em função de outro.

|  |  |  |
| --- | --- | --- |
|  |  | (16) |

Segundo Montgomery et al (2003), reescrevendo a equação (14) em termos de , temos que que é:

|  |  |  |
| --- | --- | --- |
|  |  | (17) |

Tomando a esperança dos termos e reescrevendo a equação, Montgomery et al (2003) explicam que um estimador não tendencioso para é:

|  |  |  |
| --- | --- | --- |
|  |  | (18) |

**2.1.5 Testes de hipótese e análise de resíduos**

**2.1.5.1 Teste de significância (Teste F)**

Conforme descrito anteriormente, Sharpe et al (2003) explicitam a necessidade de verificar as adequações do modelo. O teste de significância, ou teste F, verifica se existe uma relação linear entre a variável resposta e o as variáveis independentes repressoras. O teste de significância envolve o cálculo de um valor-p, que é a probabilidade de obter um resultado igual ou mais extremo do que o observado, assumindo que a hipótese nula é verdadeira. Se o valor-p for menor do que um nível de significância pré-determinado a hipótese nula é rejeitada, indicando que há evidências estatísticas suficientes para suportar a hipótese alternativa de que há uma relação linear significativa entre as variáveis independentes e dependentes. As hipóteses são:

|  |  |  |
| --- | --- | --- |
|  |  | (19) |

A estatística F é descrita por Montgomery et al (2003) como:

|  |  |  |
| --- | --- | --- |
|  |  | (20) |

Rejeita-se quando:

|  |  |  |
| --- | --- | --- |
|  |  | (21) |

**2.1.5.2 Análise de resíduos**

A análise de resíduos é uma etapa crucial para avaliar a qualidade de ajuste de um modelo de regressão, segundo Pelli Neto (2003). Para Sharpe et al (2011) é necessário avaliar a homoscedasticidade, normalidade e não-correlação dos resíduos. A correlação entre os erros pode indicar que o modelo de regressão não é adequado para representar a relação de dependência entre as variáveis. A análise gráfica dos dados e dos resíduos pode ajudar a detectar essas características, especialmente em dados coletados ao longo do tempo.

São métodos e testes amplamente utilizados para análise de resíduos:

* Gráfico Q-Q Plot para verificação de normalidade.
* Segundo Razali & Wah (2011), para dar suporte aos métodos gráficos, métodos mais formais e numéricos devem ser realizados antes de tirar qualquer conclusão sobre a normalidade dos sendo opções os testes de Anderson-Darling, Shapiro-Wilk, Kolmogorov-Smirnov e Lilliefors.
* Para teste da independência dos resíduos, o teste de Durbin-Watson deve ser utilizado, caso a independência seja violada, o modelo não representa de forma adequada a linearidade dos dados.
* Gráfico de resíduos x valores ajustados para verificar homoscedasticidade.
* O pressuposto da homoscedasticidade dos resíduos também pode ser verificado pelos testes de Breusch-Pagan ou Goldfeld-Quandt.

Existem outros testes que podem ser utilizados para verificar a qualidade do ajuste do modelo em relação aos resíduos produzidos. Os descritos anteriormente são os mais utilizados pela literatura e que já podem detectar sua adequação ou não.

**2.2 Modelos Lineares Generalizados e regressão não-linear**

Modelos Lineares Generalizados (MLG) são uma extensão dos modelos lineares de regressão múltipla que foram propostos por Nelder e Wedderburn (1972). Abrangem uma ampla gama de modelos estatísticos aplicáveis em diversas áreas de pesquisa. Os MLG apresentam vantagens em relação aos modelos clássicos, pois permitem a modelagem de variáveis resposta que seguem a família exponencial de distribuições, como por exemplo, distribuição binomial, Poisson e gama.

Outro ponto é que os modelos lineares generalizados fornecem maior flexibilidade na relação funcional entre a média da variável resposta e o preditor linear. Isso é possível devido à introdução de uma função de ligação que pode ser escolhida de acordo com as características do problema em questão. Essa função não precisa ser necessariamente a identidade e pode assumir qualquer forma monótona não-linear.

Os MLG estão dentro de uma classe notável dentro da estatística, a família exponencial. A família exponencial de distribuições é uma classe paramétrica que engloba muitas distribuições conhecidas, como a distribuição normal, binomial, binomial negativa, gama, Poisson, normal inversa, multinominal, beta e logarítmica, entre outras. Essa família de distribuições apresenta propriedades estatísticas importantes, como a propriedade de suficiência estatística, que a tornam amplamente utilizada na teoria estatística.

Na análise de modelos de regressão linear, é comum utilizar distribuições dentro da família exponencial como distribuições para a variável resposta. Isso ocorre devido às propriedades matemáticas convenientes dessas distribuições e à flexibilidade que elas oferecem para modelar a relação entre a média da variável resposta e os preditores.

É importante ressaltar que a família exponencial de distribuições não é restrita apenas a modelos de regressão linear, mas também pode ser utilizada em outras áreas da estatística, como em análise de sobrevivência e modelos de contagem.

Os modelos lineares da família exponencial podem ser escritos de forma generalizada como:

|  |  |  |
| --- | --- | --- |
|  |  | (22) |

Onde, são conhecidos, é o parâmetro de disperão conhecido e o parâmetro canônico. Segundo Costa (2019), está associado unicamente à variância da resposta. Segundo o autor, são propriedades da família exponencial:

|  |  |  |
| --- | --- | --- |
|  |  | (23) |

Segundo Costa (2019) algoritmo para estimação dos MLG foi proposto por McCullagh e Nelder (1989), e tem objetivo de maximizar a função de verossimilhança. Para análise da qualidade dos ajustes, observa-se a minimização da função Deviance na forma:

|  |  |  |
| --- | --- | --- |
|  |  | (24) |

As funções de log-verossimilhança saturada e log-verossimilhança se modificam de acordo com as distribuições como mostra a tabela abaixo:

Table 1 – Funções desvios para alguns modelos

Text, letter

Description automatically generated

Fonte: Cordeiro (2008)

Para Costa (2019), é usual a comparação do valor observado da função com a distribuição Qui-Quadrado com graus de liberdade, onde é o tamanho da amostra e k é o número de parâmetros do modelo (incluindo intercepto). Dessa forma o autor afirma que assintoticamente a estatística D sob hipótese de o modelo é compatível com o comportamento esperado dos dados, é comparado a distribuição Qui-Quadrado com graus de liberdade, na forma:

|  |  |  |
| --- | --- | --- |
|  |  | (25) |

Para Sousa (2010), em termos gerais, há superdispersão dos dados em relação ao modelo ajustado, ou seja, a variância esperada é maior que a prevista para o modelo, quando o valor-p é 0. A subdispersão é o fenômeno contrário ao da superdispersão e ocorre quando o valor-p é igual a 1. As soluções para os fenômenos citados é a procura de outro modelo que melhor se encaixe na distribuição de dados ou realizar a estimação do parâmetro no intervalo , para ajustar a qualidade do modelo.

**2.2.1 Modelo Gama**

Dentro da família exponencial, o modelo gama é usado nas situações em que a variância do modelo é proporcional ao quadrado da média (

A distribuição Gama tem a seguinte densidade de probabilidade:

|  |  |  |
| --- | --- | --- |
|  |  | (26) |

Onde é o parâmetro de dispersão e é definido por:

|  |  |  |
| --- | --- | --- |
|  |  | (27) |

As propriedades da distribuição Gama são

|  |  |  |
| --- | --- | --- |
|  |  | (28) |

**2.3 Bootstrap**

Bootstrap é uma técnica estatística desenvolvida por Efron (1979) que consiste em reamostragem aleatória para estimar a distribuição de um conjunto de dados. Inicialmente o método foi desenvolvido como uma alternativa a outra técnica estatística de reamostragem (Jackknife) e sua aplicação se mostrou mais ampla e aplicável.

Segundo o autor, dada uma variável aleatória com distribuição desconhecida e, deseja-se estimar a distribuição dos dados a partir da amostra. É assumido que a amostra pode ser representada como uma pseudopopulação, com características da verdadeira população. Por meio da geração repetida de amostras aleatórias (reamostras) desta pseudopopulação, a distribuição de amostragem de uma estatística pode ser estimada. Dessa forma, o procedimento descrito anteriormente pode ser considerado como Bootstrap não-paramétrico, já que não é necessário o conhecimento da distribuição dos dados *a priori.* Da Cunha et al (2003) descrevem em seu trabalho a utilização da técnica de Bootstrap paramétrico, procedimento que tem vasta utilização em diversos campos de estudo. Entretanto, o tema não será abordado neste trabalho.

Abaixo a Figura 1 descreve o procedimento Bootstrap:

Figura 1 - Procedimento Bootstrap

Diagram

Description automatically generated

Fonte: Autor

**2.2.1 Intervalo de confiança percentílico para Bootstrap**

Segundo Neyman (1937) Intervalo de confiança (IC) se refere a um intervalo de valores, calculado a partir de uma amostra aleatória, dentro do qual é provável que o valor de um parâmetro populacional desconhecido esteja contido com um certo nível de confiança. Esse nível de confiança é geralmente expresso em termos de uma porcentagem, como 95%, e indica a probabilidade de que o intervalo de confiança contenha o verdadeiro valor do parâmetro.

Autores como Hung et al (2014) afirmaram que o intervalo de confiança pode ser usado para estimar o grau de incerteza associado à estimativa do parâmetro e fornecer informações sobre a precisão da estimativa. Elaborando, os intervalos de confiança são importantes para a estatística porque fornecem uma medida da incerteza associada a uma estimativa de parâmetro populacional a partir de uma amostra aleatória. Essa medida de incerteza é crítica para interpretar os resultados de estudos estatísticos e para tomar decisões baseadas em evidências estatísticas.

Para a técnica de Bootstrap não-paramétrico é possível utilizar o intervalo de confiança percentílico, que segundo Franco et al (2005) pode ser descrito da seguinte forma:

1. É gerado uma quantidade de amostras de bootstrap a partir dos dados observados
2. Em seguida o parâmetro de estatístico de interesse estimado em cada uma dessas reamostras Bootstrap é calculado
3. Os valores do parâmetro estatístico calculados são ordenados em ordem crescente
4. Os percentis de interesse para construir o intervalo de confiança são selecionados (.
5. Podemos definir os limites interior e superior de um intervalo de confiança Bootstrap percentílico ao nível de (1- como sendo:

|  |  |  |
| --- | --- | --- |
|  |  | (23) |

**2.4 Programação Linear**

Dantzig (2002) descreva a programação linear como uma técnica matemática utilizada para maximizar ou minimizar uma função linear sujeita a um conjunto de restrições lineares. Essa técnica é amplamente utilizada em problemas de otimização em áreas como economia, engenharia, logística, entre outras. Para Baio et al (2004), programação linear é uma técnica para o planejamento de atividades para alcançar um resultado ideal, levando em consideração as opções possíveis. Em um problema de programação linear, normalmente existem diversas soluções possíveis, desde que todas as restrições sejam cumpridas, embora haja apenas uma solução ótima.  
 Frossard (2009) descreve que o objetivo do modelo linear é descobrir o valor mais adequado para uma função, mediante a imposição de um conjunto de restrições lineares de natureza estrita e não estrita. É um modelo matemático de otimização linear, que engloba variáveis de decisão, uma função-objetivo e restrições técnicas expressas por meio de inequações lineares. As inequações lineares, que representam as restrições técnicas, devem ser satisfeitas simultaneamente pelas variáveis de decisão para que a função-objetivo seja maximizada ou minimizada. O modelo linear é amplamente utilizado em diversas áreas para resolver problemas de otimização.

A formulação de um problema de programação linear pode ser escrita da seguinte forma, conforme proposta por Lewis (2008):

|  |  |  |
| --- | --- | --- |
|  |  | (24) |

As variáveis ​​ são chamados de variáveis ​​de decisão, e seus valores estão sujeitos a restrições. Um conjunto de satisfazendo todas as restrições é chamado de ponto viável e o conjunto de todos esses pontos é chamado de região viável. A solução do programa linear deve ser um ponto () na região viável, ou então nem todas as restrições seriam satisfeitas.

Aplicando a forma de álgebra linear, temos que:

|  |  |  |
| --- | --- | --- |
|  |  | (24) |

Na forma matricial:

|  |  |  |
| --- | --- | --- |
|  |  | (25) |

A matriz possui formato cuja coluna é . Esta matriz corresponde aos coeficientes em nas restrições de um problema de programação linear. O vetor é um vetor de soluções para o problema, é o vetor do lado direito e é o vetor de coeficientes.

Dantzig (1951) desenvolveu um método para resolver os problemas de programação linear. O Simplex é um algoritmo utilizado para resolver problemas de otimização em que é necessário encontrar a melhor solução entre várias opções. Ele funciona iterativamente, explorando soluções viáveis até encontrar a solução ótima, que é aquela que maximiza ou minimiza a função-objetivo sujeita às restrições do problema. O método Simplex é baseado em uma geometria simples, utilizando conceitos de simplex e operações matriciais para encontrar as soluções ótimas.

Lima Pinto et al (2008) descrevem que a ideia geométrica subjacente consiste em percorrer, passo a passo, os pontos extremos adjacente do conjunto viável em um problema de programação linear, aprimorando o valor da função objetivo a cada extremo visitado.

Atualmente o algoritmo Simplex é utilizado computacionalmente de forma ampla por vários programas relacionados a resolução de problemas que envolvem programação linear, Fearnley e Savani (2015) afirmam que o método é conhecido por performar bem na prática, apesar de existir métodos mais modernos que reduzem a complexidade computacional necessária para resolver problemas muito grandes.

**3. Metodologia**

Para a elaboração deste trabalho, optou-se pela abordagem de uma pesquisa descritiva ou quantitativa e que dispõe de um estudo de caso. Para Silva & Simon (2005) as pesquisas quantitativas são mais adequadas nas ocasiões em que se sabe as qualidades do objeto de estudo e se tem o controle do que vai se pesquisar. Para Silva & Junior (2014), os dados da abordagem quantitativa têm necessariamente natureza numérica, como grandezas monetárias, físicas ou escala de atitude, portanto é imperativo o uso de tal método.

Godoy (1995) descreve o estudo de caso como um tipo de pesquisa no qual o objeto é uma parte que se analisa de forma profunda. O autor ainda acrescenta que a estratégia do estudo de caso procura responder questões relacionada as razões e modos sobre como certos fenômenos ocorrem e no enfoque exploratório o pesquisador deve estar aberto a novas descobertas.

**4. Estudo de caso**

**4.1 Base de dados**

Conforme descrito anteriormente, a base de dados utilizada nessa pesquisa passou por processo de avaliação, crítica, auditoria e validação pelos operadores do SEB e pela sociedade. O conjunto está disponível online no portal da reguladora (<https://www.gov.br/aneel/pt-br>), com todas as informações relativas à tomada de subsídio.

A base de dados foi disponibilizada em forma de planilha eletrônica contendo 125 observações (linhas) e 20 colunas. A planilha contempla as seguintes informações conforme descrito pela NT 97/2022–SRM/ANEEL:

* Concessionária
* Tipo (holding, não-licitada)
* Ano
* Identificação dos agentes
* Custos operacionais contábeis (PMSO)
* Ativos físicos
  + Equipamentos de rede com tensionamento menor de 230 kV
  + Equipamentos de rede com tensionamento maior de 230 kV
  + Equipamentos de subestação com tensionamento menor de 230 kV
  + Equipamentos de subestação com tensionamento maior de 230 kV
  + Módulos de manobra com tensionamento menor de 230 kV
  + Módulos de manobra com tensionamento maior de 230 kV
* Potência total de equipamentos de subestação
  + Potência aparente: MVA
  + Potência reativa: MVar
* Indisponibilidade de rede (não utilizado na pesquisa)
* Idade Média (não utilizado na pesquisa)
* Adversidade (não utilizado na pesquisa)
* Resultados do modelo de DEA (4 campos não utilizados na pesquisa)

Na NT 97/2022–SRM/ANEEL a reguladora explicita que foram identificadas e agrupadas as empresas que efetivamente realizam algum tipo de compartilhamento operacional. A estratégia tem por objetivo mitigar potenciais inconsistências na alocação de custos em algumas transmissoras e Sociedades de Propósito Específico – SPE que pertencem ao mesmo grupo econômico e eventualmente compartilham custos. Esse agrupamento é importante de se ressaltar pois pode trazer distorções nos dados uma vez que, agrupamentos ou holdings artificiais trazem assimetrias entre agentes que pouco tem semelhanças em termos de custo, ativos físicos e potência de equipamentos. Outro ponto, é que a reguladora eliminou da consolidação final de dados com PMSO negativo, PMSO nulo e sem produtos. Dessa forma, na base de dados constam 28 agrupamentos de concessionárias.

**4.2 Descrição dos dados**

**4.3. Modelo linear**

**4.3.1 Modelo linear leave-one-out**

**4.4 Modelo de programação linear**

**4.5 Modelo de programação linear com aplicação de bootstrap**

**4.6 Resultados**

**4.7 Discussões**

**5. Conclusão**

Links

<https://www.abradee.org.br/setor-eletrico/visao-geral-do-setor/> <acessado em 19-11-2022, 17h18>

<https://www.isacteep.com.br/ri/a-isa-cteep/setor-de-transmissao> <acessado em 21-11-2022 ,11h42>

para dar suporte aos métodos gráficos, métodos mais formais que são os

métodos numéricos e testes formais de normalidade devem ser realizados antes de tirar qualquer conclusão sobre a normalidade dos dados

REFERÊNCIAS BIBLIOGRÁFICAS (APA)

Brasil (1996). Lei Nº 9.427, de 26 de dezembro de 1996. Institui a Agência Nacional de Energia Elétrica - ANEEL, disciplina o regime das concessões de serviços públicos de energia elétrica e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 27 dez. 1999. Disponível em: < http://www.planalto.gov.br/ccivil\_03/leis/l9427cons.htm>. Acesso em: 24 mar. 1996.

Brasil (2007). Ministério de Minas e Energia. Agência Nacional de Energia Elétrica. Superintendência de Regulação Econômica e Estudos do Mercado. Nota Técnica Nº 204/2018 – SRM/ANEEL

Brasil (2018). Ministério de Minas e Energia. Agência Nacional de Energia Elétrica. Superintendência de Regulação Econômica e Estudos do Mercado. Nota Técnica Nº 204/2018 – SRM/ANEEL.

Baio, F. H., Antuniassi, U. R., Balastreire, L. A., & Caixeta Filho, J. V. (2004). Modelo de programação linear para seleção de pulverizadores agrícolas de barras. Engenharia Agrícola, 24, 355-363.

Costa, M. A. (2019). Tópicos em ciência dos dados: introdução dos modelos paramétricos e suas aplicações utilizando o R.

Costa, M. A., Salvador, C. V. M., & da Silva, A. V. (2022). Stochastic Data Envelopment Analysis applied to the 2015 Brazilian energy distribution benchmarking model. *Decision Analytics Journal*, *3*, 100061.

Cordeiro, G. M., & Demétrio, C. G. (2008). Modelos lineares generalizados e extensões. Piracicaba: USP, 31.

da CUNHA, W. J., & COLOSIMO, E. A. (2003). Intervalos de confiança bootstrap para modelos de regressão com erros de medida. Rev. Mat. Estat, 21(2), 25-41.

Dantzig, G. B. (1951). Maximization of a linear function of variables subject to linear inequalities. Activity analysis of production and allocation, 13, 339-347.

Dantzig, G. B. (2002). Linear programming. Operations research, 50(1), 42-47.

Efron, B. (1979). Bootstrap methods: another look at the jackknife annals of statistics 7: 1–26. View Article PubMed/NCBI Google Scholar, 24.

Fearnley, J., & Savani, R. (2015, June). The complexity of the simplex method. In Proceedings of the forty-seventh annual ACM symposium on Theory of computing (pp. 201-208).

Feitosa Neto, S. D. A. (2009). Um estudo da metodologia de desconto da Receita Anual Permitida (RAP) em função dos atrasos na entrada em operação de empreendimentos de transmissão

Franco, G., Reisen, V., & Almeida, N. (2005). Intervalos De Confiança Bootstrap Para O Parâmetro D Em Modelos De Integração Fracionária. Simpósio Brasileiro de Pesquisa Operacional.37, 735-744.

Frossard, A. C. P. (2009). Programação linear: maximização de lucro e minimização de custos. Revista Científica da Faculdade Lourenço Filho, 6(1).

Galton, F. (1891). Hereditary genius. D. Appleton. Hoffmann, R., Vieira, S. (1977). Análise de regressão. Uma introdução à econometria, 2.

Godoy, A. S. (1995). Pesquisa qualitativa: tipos fundamentais. Revista de Administração de empresas, 35, 20-29.

Hung, R. K., Al-Mallah, M. H., McEvoy, J. W., Whelton, S. P., Blumenthal, R. S., Nasir, K., ... & Blaha, M. J. (2014). Prognostic value of exercise capacity in patients with coronary artery disease: the FIT (Henry Ford ExercIse Testing) project. In Mayo Clinic Proceedings (Vol. 89, No. 12, pp. 1644-1654). Elsevier.

Kennedy, P. E. (2002). Sinning in the basement: What are the rules? The ten commandments of applied econometrics. Journal of Economic Surveys, 16(4), 569-589.

Lewis-Beck, C., & Lewis-Beck, M. (2015). Applied regression: An introduction (Vol. 22). Sage publications.

Lewis, C. (2008). Linear programming: theory and applications. Whitman College Mathematics Department.

de Lima Pinto, L., & Menezes, M. A. F. (2008). Implementação de algoritmos simplex e pontos interiores para programação linear. Revista EVS-Revista de Ciências Ambientais e Saúde, 35(2), 225-246.

Lopes, A. L. M., de Almeida Vilela, B., Costa, M. A., & Lanzer, E. A. (2016). Crítica do modelo de cálculo do custo eficiente das empresas brasileiras de distribuição de energia elétrica. Revista Gestão e Tecnologia, 16(3), 5-30.

Milone, Giuseppe; Angelini, Flavio. (1995). Estatística aplicada. São Paulo: Atlas, 1995. 286 p.

Montgomery, D. C., Runger, G. C. (2003). Estatística aplicada e probabilidade para engenheiros, 2ª. Ed. Rio de Janeiro: Editora LTC, 416.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. Journal of the Royal Statistical Society: Series A (General), 135(3), 370-384.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 236(767), 333-380.

Pelli Neto, A. (2003). Curso de Engenharia de Avaliação Imobiliária–fundamentos e aplicação da estatística inferencial. Belo Horizonte, MG.

Pessanha, J. F. M., Mello, M. A. R. F. D., Barros, M., & Souza, R. C. (2010). Avaliação dos custos operacionais eficientes das empresas de transmissão do setor elétrico Brasileiro: uma proposta de adaptação do modelo DEA adotado pela ANEEL. Pesquisa Operacional, 30, 521-545.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. Journal of statistical modeling and analytics, 2(1), 21-33.

Sharpe, N. R. (2000). Estatística aplicada: administração, economia e negócios. Grupo A-Bookman.

Silva, D., Lopes, E. L., & Junior, S. S. B. (2014). Pesquisa quantitativa: elementos, paradigmas e definições. Revista de Gestão e Secretariado, 5(1), 01-18.

Silva, D. & Simon, F. O. (2005). Abordagem quantitativa de análise de dados de pesquisa: construção e validação de escala de atitude. Cadernos do CERU, 2(16), 11-27.

Sousa, K. M. D. M. (2010). Modelos lineares generalizados e modelos de dispersão aplicados à modelagem de sinistros agrícolas (Doctoral dissertation, Universidade de São Paulo).

Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. Journal of Statistics Education, 9(3).

Upton, G., & Cook, I. (2014). *A dictionary of statistics 3e*. Oxford university press.

Links

<https://antigo.aneel.gov.br/web/guest/tomadas-de-subsidios?p_p_id=participacaopublica_WAR_participacaopublicaportlet&p_p_lifecycle=2&p_p_state=normal&p_p_mode=view&p_p_cacheability=cacheLevelPage&p_p_col_id=column-2&p_p_col_count=1&_participacaopublica_WAR_participacaopublicaportlet_ideDocumento=47188&_participacaopublica_WAR_participacaopublicaportlet_tipoFaseReuniao=fase&_participacaopublica_WAR_participacaopublicaportlet_jspPage=%2Fhtml%2Fpp%2Fvisualizar.jsp#:~:text=decis%C3%B5es%20da%20Ag%C3%AAncia.-,NOTA%20T%C3%89CNICA%20N%C2%BA%2097%2F2022%E2%80%93SRM%2FANEEL%20Em,29%20de%20julho%20de%202022.&text=para%20consolida%C3%A7%C3%A3o%20da%20Base%20de,Lei%20n%C2%BA%2012.783%2C%20de%202013>.

Brasil. Ministério de Minas e Energia. Agência Nacional de Energia Elétrica.

Tomada de susidio

<https://antigo.aneel.gov.br/web/guest/tomadas-de-subsidios?p_auth=imLneOz1&p_p_id=participacaopublica_WAR_participacaopublicaportlet&p_p_lifecycle=1&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1&_participacaopublica_WAR_participacaopublicaportlet_ideParticipacaoPublica=3700&_participacaopublica_WAR_participacaopublicaportlet_javax.portlet.action=visualizarParticipacaoPublica>

Para verificar se há uma relação linear subjacente no modelo de regressão múltipla, é necessário verificar a condição de linearidade para cada um das variáveis explicativas através do diagrama de dispersão de y versus cada variável X. Assim como na regressão simples, também se sugere plotar o diagrama dos resíduos, a fim de constatar possíveis violações das condições de linearidade.

Proposta de abertura de Tomada de Subsídios para consolidação da Base de Dados que será utilizada para subsidiar o estudo de benchmarking de custos

operacionais regulatórios das concessionárias de transmissão prorrogadas nos termos da Lei nº 12.783, de 2013.