

Universidade Federal do Paraná  
Setor de Ciências Exatas  
Departamento de Estatística  
Programa de Especialização em *Data Science* e *Big Data*

Rafael Felice Fan Chen

**Estudo de modelos não lineares para estimação  
do frete rodoviário agrícola no Brasil**

**Curitiba  
2023**

Rafael Felice Fan Chen

# **Estudo de modelos não lineares para estimação do frete rodoviário agrícola no Brasil**

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Walmes Marques Zeviani

Curitiba  
2023

# Estudo de modelos não lineares para estimação do frete rodoviário agrícola no Brasil

Rafael Felice Fan Chen<sup>1</sup>, Walmes Marques Zeviani<sup>2</sup>

<sup>1</sup>Aluno do programa de Especialização em Data Science & Big Data\*

<sup>2</sup>Professor do Departamento de Estatística - DEST/UFPR<sup>†</sup>

O valor das commodities agrícolas é influenciado por diversos fatores, dentre estes fatores, o valor do frete rodoviário tem um grande impacto, podendo ser o ponto crucial entre o lucro e o prejuízo da safra para o produtor. Com isso, este trabalho tem como objetivo calibrar e validar um modelo que consiga estimar o valor do frete rodoviário e ajudar o produtor a tomar melhores decisões para a safra. O modelo utilizado foi o de regressão segmentada com restrições. Além de calibrar o modelo, a partir da metodologia aplicada, foi possível utilizando várias técnicas de ETL (Extração, Transformação e Carga), o estudo também possibilitou entender e identificar a variável que mais influencia o preço do frete rodoviário no Brasil e os meses de maior, menor e média histórica de volume de exportação, sendo o mês de Abril o de maior e Janeiro o de menor volume histórico. Foi possível também verificar que o produtor é mais prejudicado pelo valor mais alto do frete nos meses com maior volume da produção, havendo indícios que seja devido a alta na demanda de caminhões no período. De modo geral, o modelo escolhido conseguiu descrever bem a relação do preço do frete como função da distância percorrida.

**Palavras-chave:** Logística, Regressão segmentada, Commodities

## 1. Introdução

O Brasil é um dos maiores produtores e exportadores mundiais de commodities agrícolas, contribuindo significativamente para o suprimento global de alimentos e para a geração de *divisas(receitas??)*. Nos últimos 10 anos a produção brasileira de soja e milho aumentou em mais de 75%, tendo como destaque a cultura da soja, que consolidou o Brasil como o maior produtor e exportador dessa commodity. *(Referência)*

Os Estados do Mato Grosso, Paraná, Rio Grande do Sul, Goiás e Minas Gerais são onde o cultivo da cultura da soja mais se destaca no Brasil. O cultivo da soja geralmente começa entre os meses de setembro e outubro, com o plantio da safra de verão. A colheita ocorre entre os meses janeiro e março, dependendo da região. Após a colheita da safra de verão, alguns agricultores optam por realizar o plantio da segunda safra de soja, conhecida como safrinha. Nesse caso, o plantio ocorre em janeiro ou fevereiro e a colheita entre junho e julho. *(Referência)*

Quanto ao milho, o Brasil possui duas safras: a primeira safra, também conhecida como "milho verão" ou "milho de safra", e a segunda safra, chamada usualmente de "milho safrinha". A primeira safra de milho é plantada entre setembro e outubro e colhida entre janeiro e março, essa safra é cultivada principalmente nos estados do Paraná, Mato Grosso do Sul, Goiás e Minas Gerais. Após a colheita da primeira safra, é realizado o plantio do milho safrinha, que ocorre entre janeiro e fevereiro e é colhido entre maio e julho. O milho safrinha é cultivado principalmente nos estados do Mato Grosso, Mato Grosso do Sul e Goiás. *(Referência)*

Para a safra 22/23 no Brasil, se estima que a produção de soja e milho atinjam o volume recorde de 288 milhões de toneladas e que dessa produção 148 milhões de toneladas sejam para exportação (USDA, 2022).

Neste contexto, um transporte eficiente, confiável e com preço justo é fundamental para o sucesso desse setor, pois a produção se concentra longes dos portos de exportação o que dificulta a sua escoação. Assim entra o estudo da logística que envolve desde a gestão da distribuição, do manuseio de cargas e desempenha um papel crucial na agricultura brasileira, especialmente

\*rafaelchen@agrinvest.agr.br

†walmes@ufpr.br

quando se trata do transporte para a exportação das commodities agrícolas. (*Referência*)

O modal mais usado para escoar a produção agrícola no Brasil é o rodoviário. E verifica que os fatores que mais influenciam o preço do frete agrícola são o tamanho médio da carga transportada, distância média do frete, custos monetários, seguro da carga, preço do combustível, a possibilidade do frete retorno, aspectos geográficos, tempo de carga e descarga (CAIXETA-FILHO, 2003). Porém a distância percorrida foi o fator fundamental para a composição dos preços dos fretes praticados (APARECIDA et al., 2013).

Outro aspecto importante do frete para o agronegócio é o seu impacto na formação do preço das commodities. O Preço da Paridade de Exportação (PPE) de uma commodity leva em consideração o preço negociado na bolsa de valores, os bônus e as despesas portuárias, além do desconto referente à logística até o ponto de referência (ARAUJO, 2017). O PPE é o preço em que um produto deve ser vendido no mercado externo para que seja possível igualar seu custo de produção e garantir a competitividade no comércio internacional (ECONOMIA, 2021).

Assim, pelo desafio enfrentado e a necessidade que o Brasil tem por ser um importante produtor e exportador de commodities agrícolas (MACHADO P. A., 2018), este artigo tem como objetivo geral, estimar o valor do frete rodoviário de grãos por quilometro rodado para os meses de maior, menor e médio histórico do volume de exportação.

## 2. Referencial teórico

### 2.1. Logística e Modal

A logística tem sua origem datada para o período segunda guerra mundial, onde foi utilizada pelos militares para controle de seus suprimentos. Hoje em dia, esta é uma importante área que se responsabiliza pelo armazenamento e transporte de mercadorias e insumos das empresas (SOUZA, 2002), inclusive de commodities agrícolas.

A logística é o processo em que visa ter um gerenciamento estratégico de compra, do transporte e no armazenamento de matérias primas, partes e produtos finais por parte da organização e de seus canais de marketing, de tal modo que a lucratividade atual e futura sejam melhor aproveitadas diante da entrega de encomendas com o menor custo possível dos produtos (CHRISTOPHER, 2007).

O Brasil é um dos países com a maior malha rodoviária mundial. Desta forma, BALLOU (2001) destaca que a seleção de um modal de transporte pode ser usada para criar uma vantagem competitiva do serviço para maximizar o lucro de uma empresa.

Os modais são classificados em cinco modalidades: aéreo, marítimo, rodoviário, ferroviário e dutoviário, cada qual com várias características, vantagens e desvantagens, quanto a capacidade, custo, velocidade, segurança, entre outros (SILVA; CURI, 2017). No Brasil, o principal modal utilizado para a escoar a produção agrícola é o rodoviário.

O modal rodoviário no Brasil dentre os demais, é o mais utilizado para transporte de passageiros e cargas. Esse modal conta com veículos que variam de pequeno à grande porte em estradas. Este modal utiliza estradas e rodovias para o transporte de pessoas e cargas e é amplamente utilizado para transporte de curta e média distância, oferecendo flexibilidade e acessibilidade (RODRIGUES; FIGUEIREDO, 2014).

## 3. Materiais e Métodos

A base de dados utilizada foram arquivos em extensão PDF enviados semanalmente por uma empresa no período de maio de 2022 até maio 2023. Nesses arquivos continham os valores praticados na semana em reais por tonelada, das principais rotas rodoviárias de movimentação de commodities agrícolas no país. Ao todo foram obtidas 394 rotas de movimentação de grãos.

O processo de ETL (Extração, Transformação e Carga) é muito difundido em locais com muitos dados, onde se encontram em diferentes fontes e que seja necessário integrá-los (SERRANO; RIQUELME, 2014). Geralmente, os passos para a aplicação do ETL são, primeiramente, a extração dos dados de interesse podendo encontrar em banco de dados, arquivos CSVs, PDFs, APIs, dentre outros. Após esse processo é realizada a transformação dos dados, para que haja uma forma de integração entre as fontes e, após isso, é realizada a carga que é onde esse novo banco é salvo, para que seja utilizado de forma inteligente e ajudar em tomada de decisões (KIMBALL; ROSS, 2011). É relevante salientar que em cada projeto pode ser realizado as etapas de acordo com seu contexto, com isso, é necessário adaptar cada abordagem do ETL para a necessidade de cada estudo.

Para organização, resumo e interpretação de conjuntos de dados, é muito utilizada a estatística descritiva que oferece uma compreensão resumida e visual

dos dados analisados, revelando padrões e características gerais dos dados coletados (MORETTIN; BUSSAB, 1997).

Foi utilizado o pacote tabulizer (LEEPER, 2022) no software R (TEAM, 2022) para a extração, tratamento e limpeza dos dados, onde, os dados semanais foram resumidos para médias mensais. Para definir quais meses seriam utilizados na análise, foi utilizado a base de dados disponibilizada pelo Ministério de Desenvolvimento e Comercio Exterior (MDIC, 2023) em que consta a média mensal de exportação de soja e milho entre os anos de janeiro de 2012 até maio de 2023. Para determinar as distâncias entre as rotas de origem e destino da soja e milho, foi utilizado o site google maps (GOOGLE, 2023), depois disso foi criado um arquivo CSV com todas as variáveis que iriam ser utilizados no presente estudo.

Um modelo estatístico é aplicado em diferentes contextos e áreas, utilizando diferentes técnicas estatísticas, como regressão linear, análise de variância, modelos de séries temporais, modelos de probabilidade, entre outros. O modelo estatístico é uma representação matemática da relação entre variáveis e tem como objetivo fazer inferências, relações e previsões, a partir de uma base de dados observados (MONTGOMERY; PECK; VINING, 2012).

Para estimar o valor do frete com base na distância percorrida, foi utilizada a técnica estatística de regressão linear segmentada, também conhecida como regressão por partes. O modelo de regressão segmentado é útil quando a relação entre as variáveis não é linear e pode variar ao longo do intervalo de dados (FRIEDMAN; MEULMAN, 2003). Essa abordagem permite ajustar vários modelos lineares aos dados, cada um correspondendo a um intervalo específico da variável independente (distância). Os pontos de quebra determinam os valores de distância onde ocorrem mudanças na inclinação da função linear. O valor exato dos pontos de quebra pode ou não ser conhecido antes da análise e, portanto, precisa ser estimado a partir dos dados (VMR, 2008). O pacote utilizado foi o segmented do R.

Para ajustar o modelo de acordo com a regra de negócio, foi utilizado o modelo de regressão segmentada com restrições (Modelo bilinear), com a restrição que as taxas sejam não negativa. Foi utilizada a seguinte equação:

$$y = \begin{cases} \theta_0 + \theta_1 \cdot x, & \text{se } x \leq \theta_b \\ \theta_0 + \theta_1 \cdot \theta_b + \theta_2 \cdot (x - \theta_b), & \text{se } x > \theta_b \end{cases}$$

onde:

$y$ : Valor do frete por tonelada;

$x$ : A distância percorrida (mil km);

$\theta_0$ : Interecepto é o valor do frete para 0 km de deslocamento;

$\theta_1$ : Taxa para cada 1000km antes do ponto de quebra;

$\theta_2$ : Taxa para cada 1000km depois do ponto de quebra;

$\theta_b$ : Ponto de quebra.

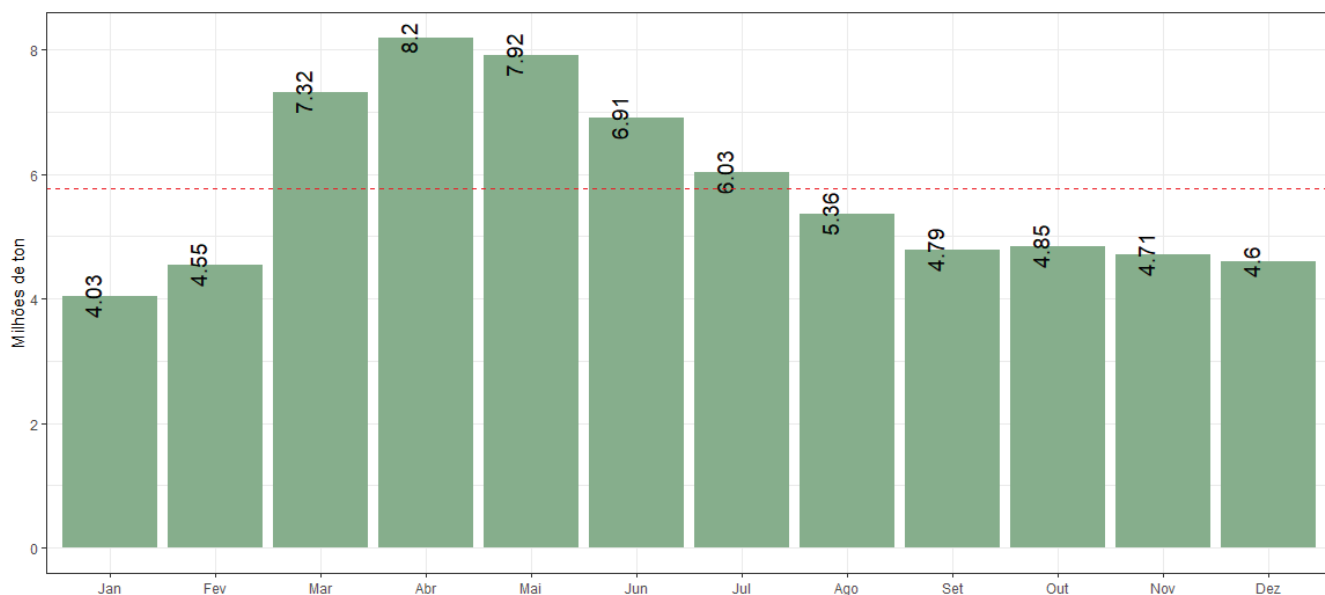
O método utilizado para estimar os parâmetros foi a máxima verossimilhança dos dados observados. A estimação por máxima verossimilhança (EVM) é uma técnica amplamente utilizada para estimação de parâmetros de um modelo estatístico que tem como principal característica encontrar valores que maximizam a verossimilhança dos dados observados (CASELLA; BERGER, 2002). Para a aplicação da técnica foi utilizado a função mle2 do pacote bbmle no R (BOLKER; TEAM., 2022), o método para incluir as restrições do modelo foi o L-BFGS-B com valores  $\theta_1 \geq 0$  e  $\theta_2 \geq 0$ .

O conceito de  $R^2$  (R-quadrado) ou coeficiente de determinação é uma métrica estatística que quantifica a qualidade do ajuste de um modelo de regressão considerando os dados observados. O valores dessa medida varia entre 0 e 1, onde quanto mais próximo de 1 indica um bom ajuste do modelo aos dados utilizados para a modelagem e valores próximos de zero indicam que o modelo não consegue explicar a variabilidade dos dados e que tem um ajuste inadequado ao dados (HAIR et al., 2006).

A taxa de incremento é uma medida que mostra a variação percentual entre dois valores, mostrando o aumento relativo entre eles. Essa medida pode ser negativa ou positiva, o seu cálculo é feito pela diferença entre dois valores e dividido pelo valor inicial, com o resultado, multiplica se por 100 para ter o resultado em porcentagem (ALBRIGHT; WINSTON; ZAPPE, 2016).

## 4. Resultados

A média mensal de exportação de soja e milho nos últimos 10 anos foi de 5,77 milhões de toneladas, sendo o mês de abril e janeiro o maior e menor meses de volume de exportação e o mês de julho o mais próximo a média histórica (Figura 1). Os meses de março, abril, maio, junho e julho tiveram suas médias acima da média histórica e o meses de janeiro, fevereiro, agosto, setembro, outubro, novembro e dezembro tiveram seu volume de exportação menor que a média histórica. A



**Figura 1:** Média mensal de exportação de soja e milho no Brasil

diferença entre a maior e menor média mensal é de mais de 4,1 milhões de toneladas.

A menor rota observada foi 241km, entre as cidades de Ponta Grossa a Paranaguá no estado do Paraná e a maior rota entre as cidades foi de Paragominas no Pará com destino ao porto de Santos em São Paulo com uma distância de aproximadamente 2.696km.

Os três meses avaliados apresentaram comportamentos bem distintos. A sazonalidade da exportação, principal destino das commodities produzidas, resulta em flutuações nos fretes. Há várias razões que confirmam esse fato, sendo os picos de safra um dos principais motivos que inflacionam o mercado. Soja e milho em grãos são exemplos de cargas que geram uma corrida no sistema de transporte de carga, pois sua movimentação aumenta consideravelmente em um determinado período do ano. O pico mais significativo ocorre durante o escoamento da soja, entre os meses de março e junho, o que frequentemente leva ao racionamento da oferta de caminhões (SOARES; F, 1997).

O intercepto, que é o coeficiente que representa o valor quando a variável "quilometro" é igual a zero, va-

riou de 4,85 a 71,75. Isso indica um custo inicial menor no mês de menor volume de exportação e tem um custo maior no mês de maior volume de exportação, que está relacionado a demanda de caminhões.

O ponto de quebra, ponto onde há uma mudança na relação entre as variáveis quilometro rodado e preço do frete, foi similar nos meses de abril 2023 e julho 2022 e menor para janeiro de 2023. Isso pode ser explicado pela origem da produção. Muda a formula de precificação e sua taxa de incremento.

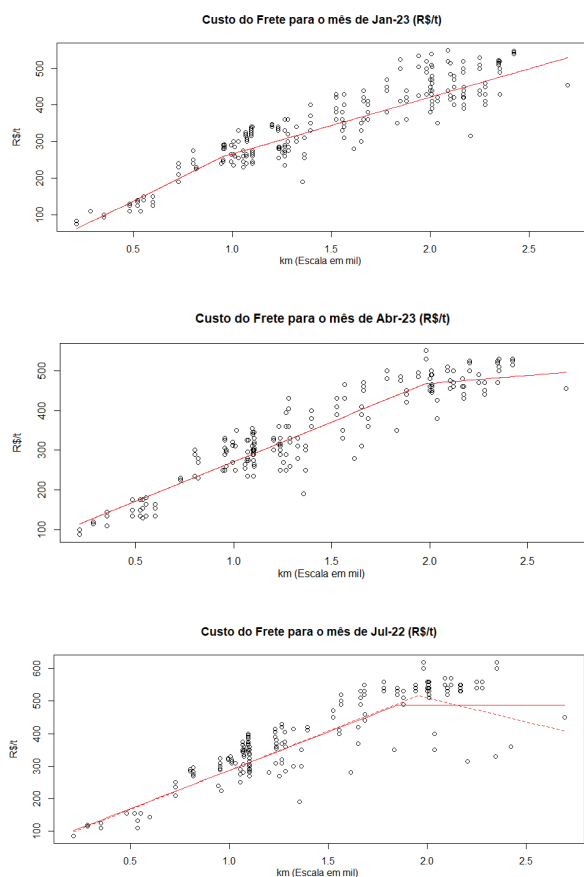
Ao utilizar o pacote segmented (MUGGEO, 2008), houve uma quebra da regra de negócio, no segundo segmento de reta o valor do quilometro decaia, isso vai contra o que é praticado no mercado, onde o custo do frete aumenta em relação ao quilometro rodado, com isso foi necessário utilizar um modelo com restrições.

O modelo segmentado bilinear obteve bons ajustes, confirmados pelos valores do  $R^2$ , assim pode se assumir que os coeficientes que o modelo apresentam tem peso considerável para análise e tem significância positiva na formação do preço do frete. Todos os valores do R-quadrado foram maiores que 0,7, indicando que os

**Tabela 1:** Estimativas dos parâmetros do modelo bilinear com restrição para os meses de Janeiro, Abril e Julho.

Mês	Intercepto( $\theta_0$ )	Ponto de quebra( $\theta_b$ )	Taxa		$R^2$
			Antes( $\theta_1$ )	Depois( $\theta_2$ )	
Janeiro-23	4,854	0,960	266,061	157,701	0,831
Abril-23	71,75	1,982	199,691	39,716	0,837
Julho-22	40,483	1,783	247,476	0,000	0,716

dados observados (quilômetro rodado) pode explicar uma grande parte da variabilidade do preço do frete. Conforme a Tabela 1, pode se verificar que os valores do R-quadrado para o mês de Abril de 2023 e Janeiro de 2023 foram praticamente idênticos, sendo 0,837 e 0,831, respectivamente. O mês de Julho de 2023 foi o mês com menor valor do ajuste, seu valor foi de 0,716.



**Figura 2:** Custo do frete para os meses de Janeiro e Abril de 2023 e Julho de 2022 por distância em quilômetros

O custo do frete em todos os meses analisados possuem o mesmo comportamento, uma tendência de aumento do valor do frete mais contundente de acordo com quilômetro rodado até o ponto de quebra. Após o ponto de quebra, o acréscimo no preço do frete é menos contundente, como podemos ver nos gráficos de dispersão na Figura 2, entretanto, o valor do frete tem uma tendência de aumentar de acordo com a distância que a produção precisa percorrer, corroborando com todos os resultados apresentados no estudo.

## 5. Conclusão

O modelo bilinear descreveu bem a relação do preço do frete como função da distância percorrida subsidiando a compreensão do fenômeno e tomada de decisão com um modelo interpretável de poucos parâmetros, obtendo valores do  $R^2$  maiores que 0,7.

Os meses que mais possui volume de exportação, necessitando assim um maior escoamento é o mês de abril e o de menor volume é o mês de Janeiro.

Em períodos de elevado volume de exportação, o ponto de ruptura logística apresentou-se mais elevado. Tal fenômeno pode ser atribuído à demanda generalizada por frete em todo o país. Por outro lado, em meses de baixo volume de exportação, a produção a ser escoada concentra-se principalmente na região Centro-Oeste do Brasil, e conseqüentemente a taxa de crescimento após o ponto de ruptura foi a maior para esse mês.

## Agradecimentos

À Agrinvest Commodities pela disponibilização dos dados e tempo para realização da pós graduação. E ao meu orientador por todos os ensinamentos.

## Referências

- ALBRIGHT, S. C.; WINSTON, W. L.; ZAPPE, C. J. Data analysis and decision making. cengage learning. 2016.
- APARECIDA, S. et al. Determinantes do preço do frete rodoviário para transporte de soja em grãos em mato grosso impactos economicos das políticas aplicadas ao agronegócio sobre o bem-estar e a produção setorial brasileiras: Uma modelagem dinâmica regionalizada. 2013.
- ARAUJO, M. O segredo do grão. 2017.
- BALLOU, R. H. Gerenciando a cadeia de suprimentos: planejamento, organização e logística empresarial. 2001.
- BOLKER, B.; TEAM., R. D. C. bbml: Tools for general maximum likelihood estimation. 2022.
- CAIXETA-FILHO, J. V. Principais determinantes do preço do frete rodoviário para transporte de soja em grãos em diferentes regiões brasileiras: uma análise econométrica assessing agrologistic risks and costs in brazil view project evaluation of public policies for the development of green corridors to agro-logistics in brazil (fapesp-belmont forum). 2003.
- CASELLA, G.; BERGER, R. L. Statistical inference. cengage learning. 2002.
- CHRISTOPHER, M. Logística e gerenciamento da cadeia de suprimentos: criando redes que agregam valor. 2007.
- ECONOMIA, M. da. Cartilha de comércio exterior: Preço de exportação. Brasília: Ministério da economia. 2021.

- FRIEDMAN, J.; MEULMAN, J. J. Multiple additive regression trees with application in epidemiology. *statistics in medicine*. 2003.
- GOOGLE. Google maps. <https://www.google.com/maps>. 2023.
- HAIR, J. J. et al. *Segmented: An r package to fit regression models with broken-line relationships*. 2006.
- KIMBALL, R.; ROSS, M. *The data warehouse toolkit: the complete guide to dimensional modeling*. [S.l.]: John Wiley & Sons, 2011.
- LEEPER, T. J. s. *tabulizer: Bindings for tabula pdf table extractor library*. 2022.
- MACHADO P. A., e. a. Logística no agronegócio: uma análise da infraestrutura de escoamento de grãos no brasil. *revista de economia agrícola*. 2018.
- MDIC. *Dados estatísticos do comércio exterior brasileiro*. 2023.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introdução ao controle estatístico da qualidade*. 2012.
- MORETTIN, P. A.; BUSSAB, W. O. *Estatística básica*. [S.l.]: Saraiva Educação SA, 1997.
- MUGGEO, V. *Segmented: An r package to fit regression models with broken-line relationships*. 2008.
- RODRIGUES, R. B.; FIGUEIREDO, K. F. Transporte rodoviário de cargas no brasil: evolução, características e desafios. *revista de economia e sociologia rural*, 5. 2014.
- SERRANO, M.; RIQUELME, J. C. Projeto e avaliação de processos etl para carregamento de data warehouse. *revista de engenharia de informação*. 2014.
- SILVA, R.; CURI, M. Custos logísticos: Um estudo sobre a composição do frete rodoviário entre zonas aduaneiras. 2017.
- SOARES, M. G.; F, J. V. C. Caracterização do mercado de fretes rodoviários para produtos agrícolas. 1997.
- SOUZA, P. T. Logística interna para empresas prestadoras de serviço. 2002.
- TEAM, R. C. R: A language and environment for statistical computing. *r foundation for statistical computing*. 2022.
- USDA. Soybean transportation guide: Brazil 2021. 2022.
- VMR, M. *segmented: an r package to fit regression models with broken-line relationships*. *r news*. 2008.