



Previsão dos fretes rodoviários da soja: uma aplicação de uma regressão múltipla

Palavras-chave: produtos agrícolas; logística agroindustrial; custo de transporte

Autores/as:

João Gabriel Duarte[Unicamp]

Andrea Leda Ramos de Oliveira[Unicamp]

Resumo

O custeamento do transporte rodoviário é uma problemática que afeta a todos na cadeia logística, pois devido a variáveis como a elevada quantidade de praças de pedágio, a instabilidade nos preços do diesel, o período do ano e distância percorrida. Sendo assim, o objetivo da pesquisa é desenvolver um modelo de predição do frete rodoviário da soja, a partir da aplicação de um modelo de regressão linear múltipla. A partir do modelo de regressão o intuito é analisar a relação funcional entre duas ou mais variáveis, tais quais assumem valores qualitativos ou quantitativos, de forma que elas possam ser preditas a partir de outras.

1. Introdução

O frete agrícola enfrenta desafios devido às diversas variáveis que influenciam diretamente a sua formação, como por exemplo: a elevada quantidade de pedágios, a instabilidade nos preços do combustível, o período do ano e a distância percorrida. O projeto em questão, teve como foco prever o valor dos fretes rodoviários da soja a partir do método de regressão múltipla.

A finalidade dos modelos de regressão é analisar a relação funcional entre duas ou mais variáveis, as quais assumem valores quantitativos ou qualitativos, de forma que elas possam ser preditas a partir de outras, ou seja, podem projetar ou estimar novos valores. Foram utilizadas variáveis como Distância, Mês, Ano, Estado Origem e Destino. Durante o ano de 2018 e 2019 foi utilizado os dados dos transportes rodoviários da soja em estados como Bahia, Distrito Federal, Goiás, Maranhão, Minas Gerais, Mato Grosso do Sul, Mato Grosso, Pará, Piauí, Paraná, Rondônia, Rio Grande do Sul, Santa Catarina, São Paulo e Tocantins.

Origem	UF	Destino	UF	Frete(R\$/t)	Distância(km)	Mês	Ano
Canarana	MT	Santos	SP	275,00	1.653	Janeiro	2019
Formosa	BA	Barreira	BA	50,6	153	Janeiro	2018
Água Boa	MT	Santos	SP	308,22	1.575	Abril	2018
Buritis	MG	Uberlândia	MG	92,5	685	Fevereiro	2018
Dom Eliseu	PA	São Luís	MA	122,75	699	Agosto	2019
Guaraí	TO	Palmeirante	TO	35,67	166	Março	2019

Quadro 1. Exemplo do banco de dados, rotas de mercado interno e exportação

2. Materiais e Métodos

O comportamento do custo rodoviário foi qualificado através da regressão de um banco de dados exemplificado pelo Quadro 1. Foram aplicados três testes até concluir quais eram as variáveis significativas para prever o custo do transporte, primeiro teste foi feito com distâncias de até 500 quilômetros e superior, e a análise se o período era de safra ou entressafra, dessa forma concluiu-se que safra e entressafra não influenciavam diretamente para a predição do frete, segunda análise foi testada distâncias de até 600 quilômetros e superior, e qual o semestre de cada transporte, entretanto a variável semestre não foi significativa, porém trocando 500 quilômetros por 600 houve um aumento no R^2 e uma diminuição do Erro (mostrado no Quadro 2), portanto foi utilizada como variável padrão as faixas de menor que 600 quilômetros e maiores ou iguais a 600 quilômetros.

Partindo da hipótese que o comportamento do frete rodoviário varia conforme distâncias menores e iguais ou superiores a 600 quilômetros, para o modal rodoviário utiliza-se:

$$Y = X\beta + \varepsilon$$

$$Y = \beta_0 + \beta_1\delta X + \beta_2(1 - \delta)X + \varepsilon$$

$$\delta_1 = \begin{cases} 0 & \text{se distância} < 600 \\ 1 & \text{se distância} \geq 600 \end{cases}$$

Em que:

$Y = \text{Frete em R\$/t}$

$X = \text{Distância em km}$

$\beta = \text{Coeficiente da regressão}$

$\delta = \text{Variável binária}$

$\varepsilon = \text{Erro aleatório}$

3. Resultados e Discussão

Variáveis	600 km	500 km	Safrá
R múltiplo	0,93745	0,93480	0,90019
R ²	0,87882	0,87386	0,81035
Erro	25,6290	26,1477	28,93715
N	2.729	2.729	2.729

Quadro 2. Resultado dos testes.

Variáveis	Coeficiente	Teste T ¹
Intercepto	31,8445	24,496
$\delta=1$	0,01055	33,365
$\delta=0$	0,01253	117,441

¹ Significativo a 0,1%.

$\delta=1$, distâncias superiores ou iguais a 600 km.

$\delta=0$, distâncias inferiores a 600 km.

R múltiplo = 0,93745

R² = 0,87873

N= 2.729

Quadro 3. Estatísticas resultantes do modal rodoviário

Observando os dados apresentados nos Quadros 2 e 3, é notório que a faixa de 600 quilômetros é a que apresenta maior R², ademais possui o menor Erro, dessa forma é a variável com maior significância dos três testes feitos. O ajuste medido pelo R² indica que 87,87% da variância da variável custo é explicada pela regressão e esses 12,13% decorrem de causas aleatórias.

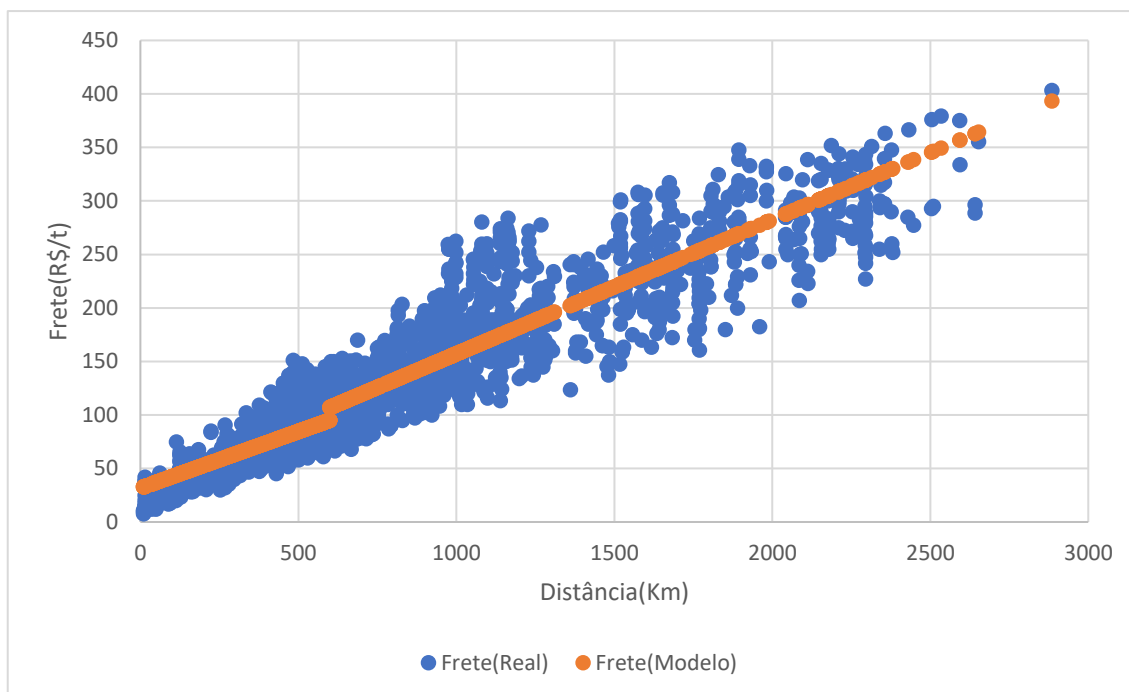


Figura 1. Gráfico mostra a dispersão do valor do frete real e o calculado pela distância

É sabido que pedágios, preço do combustível, período anual e até ano são variáveis que influenciam o custeamento do transporte rodoviário, utilizando como base a Figura 1 é possível notar que a além de apresentar um comportamento diferente para fretes até 600 quilômetros e superiores, existe uma maior competitividade no mercado para curtas distâncias.

Analisando a figura 1, apenas a dispersão na série: Frete(Modelo), é nítida a coerência e a linearidade de cada distância para com seu frete, diferente da série: Frete(Real) que apresenta total dispersão entre os pontos sem nenhuma coerência e linearidade. A faixa até 600 quilômetros apresenta uma maior competitividade no mercado e maior proximidade entre os modelos.

4. Conclusões

O modelo de previsão do frete rodoviário comparado ao frete real, trouxe maior linearidade e coerência ao custeamento, uma vez que no modelo de regressão linear múltipla o valor do frete está diretamente ligado a distância percorrida no transporte. Para transportes até 600 quilômetros, onde há maior competitividade por transportes no mercado, há também uma menor diferença entre o frete real e o frete no modelo, ou seja, o modelo de regressão múltipla para distâncias até 600 quilômetros funciona. É sabido que os períodos de safra e entressafra, não foram significativas, assim sendo a única variável significativa é a distância para o modelo.