Estimação dos principais direcionadores dos custos operacionais das empresas brasileiras de transmissão de energia elétrica utilizando modelos de regressão e programação linear

Igor Mazzeto Resende Soares
Orientador: Prof. Dr. Marcelo Azevedo Costa

Universidade Federal de Minas Gerais
Departamento de Estatística - ICEX

25 de agosto de 2023

DEST
UFMG

Agenda

DEST UFMG

Justificativa

- Importância da estimação do PMSO para determinação da RAP e consequentemente.
- Aplicação de técnicas estatísticas considerando restrições físicas e técnicas intrínsecas à operação de transmissão de energia elétrica .

DEST UFMG

Objetivos

São objetivos deste trabalho:

- Analisar as correlações existentes entre as variáveis e 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.

DEST UFMG

Metodologia

1. Pesquisa descritiva quantitativa e que dispõe de um estudo de caso.
2. Serão respondidas questões relacionadas ao funcionamento de fenômenos que occorem no âmbito do SEB.

DEST UFMG

Estudo de caso - Resultados

- Base de dados fornecidas pela NT nº 097/2022–SRM/ANEEL por meio de uma planilha eletrônica
- Base composta por 125 observações e 20 colunas
- Variável de interesse - custos operacionais compostos pelas contas de pessoal, materiais, serviços de terceiros e outros (*PMSO*)

DEST UFMG

Estudo de caso - direcionadores de custo

Descrição dos direcionadores:

1. X_1: Extensão de equipamentos de rede com tensionamento maior de 230 kV
2. X_2: Módulos de manobra com tensão igual ou superior a 230 kV
3. X_3: Equipamentos de rede com tensionamento maior de 230 kV
4. X_4: Potência aparente total,em MVA,de equipamentos de subestação
5. X_5: Potência reativa total,em Mvar,de equipamentos de subestação
6. X_6: Equipamentos de subestação com tensão inferior a 230 kV
7. X_7: Módulos de manobra com tensão inferior a 230 kV
8. X_8: Extensão de equipamentos de rede com tensionamento maior de 230 kV

DEST UFMG

Estatísticas descritivas

Tabela: Estatísticas descritivas das variáveis dependente e independentes

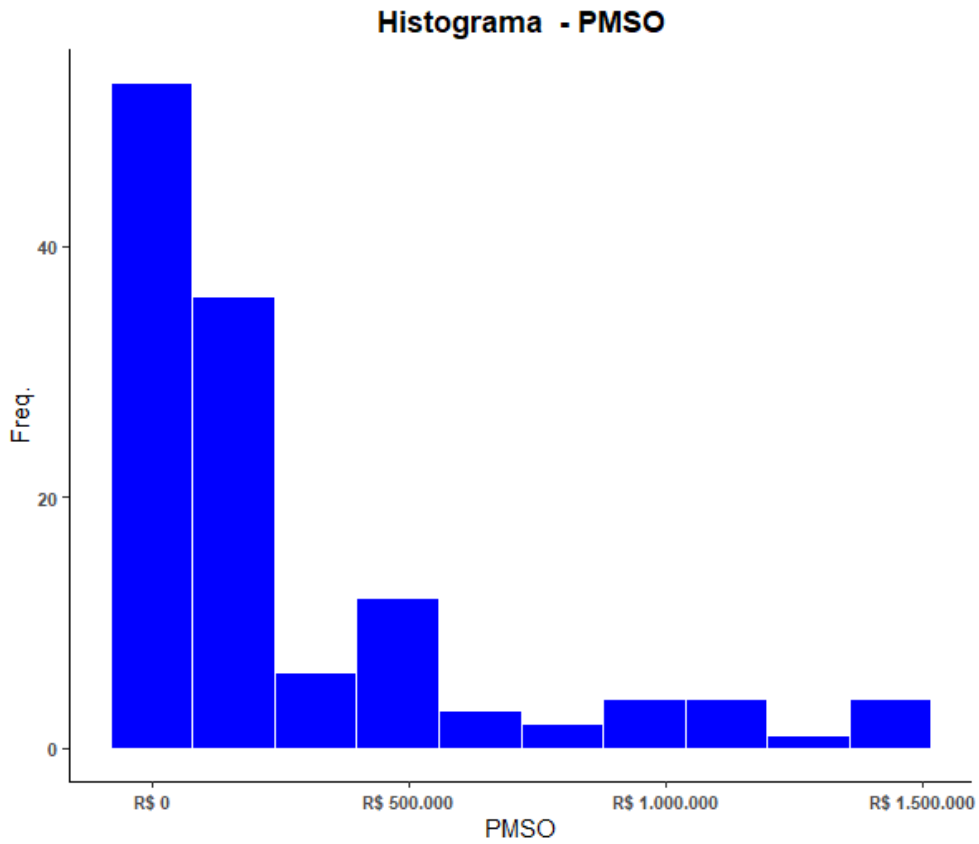
	PMSO	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Mínimo	1.273,0	0,0	4,0	7,0	0,0	0,0	0,0	0,0	0,0
1º Quartil	30.523,0	773,6	68,0	56,0	2.348,0	977,0	2,0	17,0	0,0
Mediana	103.467,0	3.517,7	138,0	116,0	7.500,0	4.227,0	5,0	43,0	45,2
Média	270.875,0	4.693,6	270,4	209,0	18.441,0	6.467,0	64,8	313,4	787,9
3º Quartil	262.355,0	6.848,9	345,0	275,0	19.527,0	9.485,0	80,0	375,0	500,7
Máximo	1.439.704,0	18.376,7	1.218,0	763,0	98.256,0	41.208,0	346,0	1.841,0	7.297,6

Fonte: Autor.

DEST UFMG

Histograma PMSO

Figura: Histograma PMSO

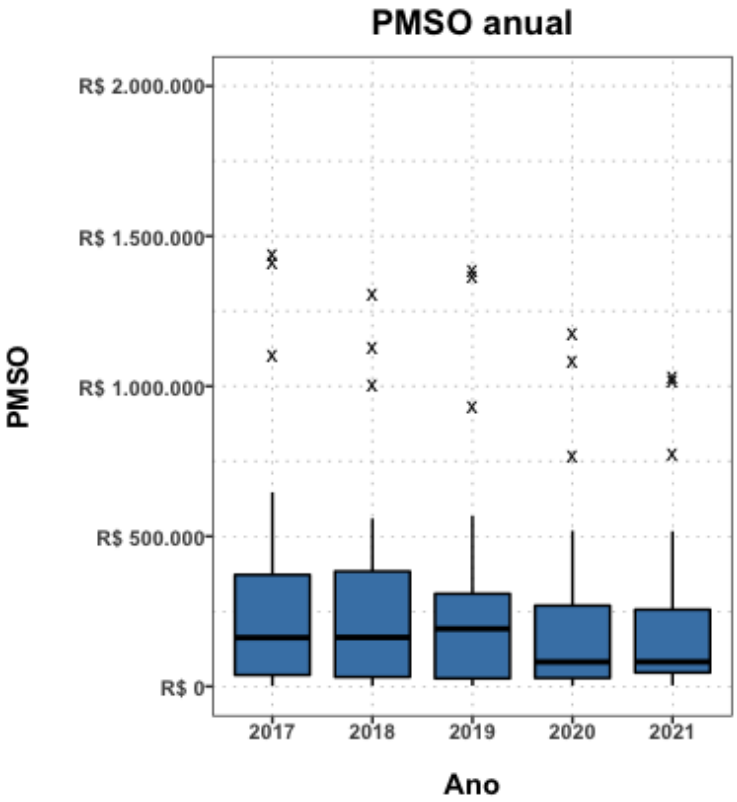


Fonte: Autor.

DEST UFMG

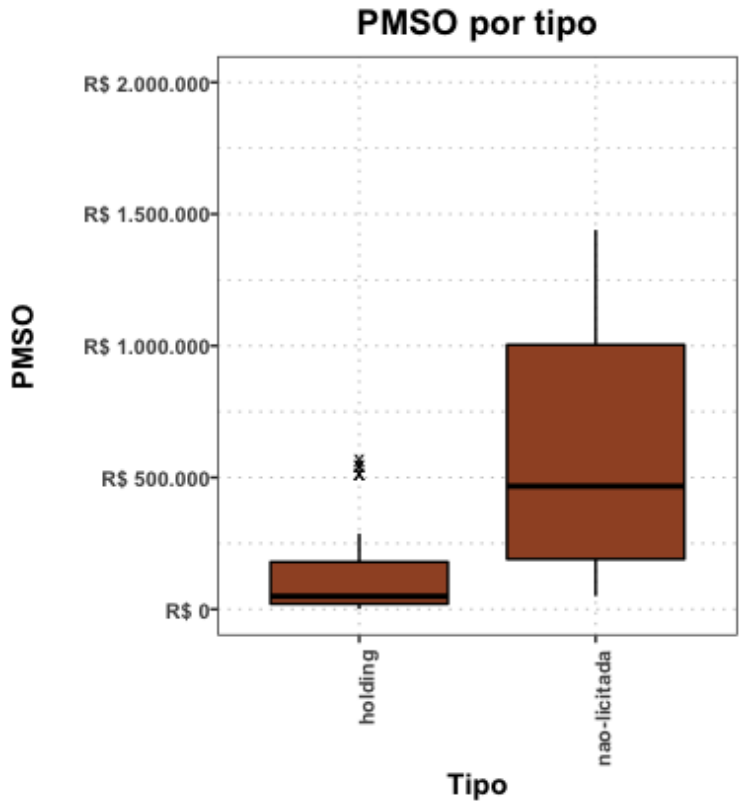
Boxplots: anual e por tipo de empresa

Figura: Boxplot anual



Fonte: Autor.

Figura: Boxplot por tipo

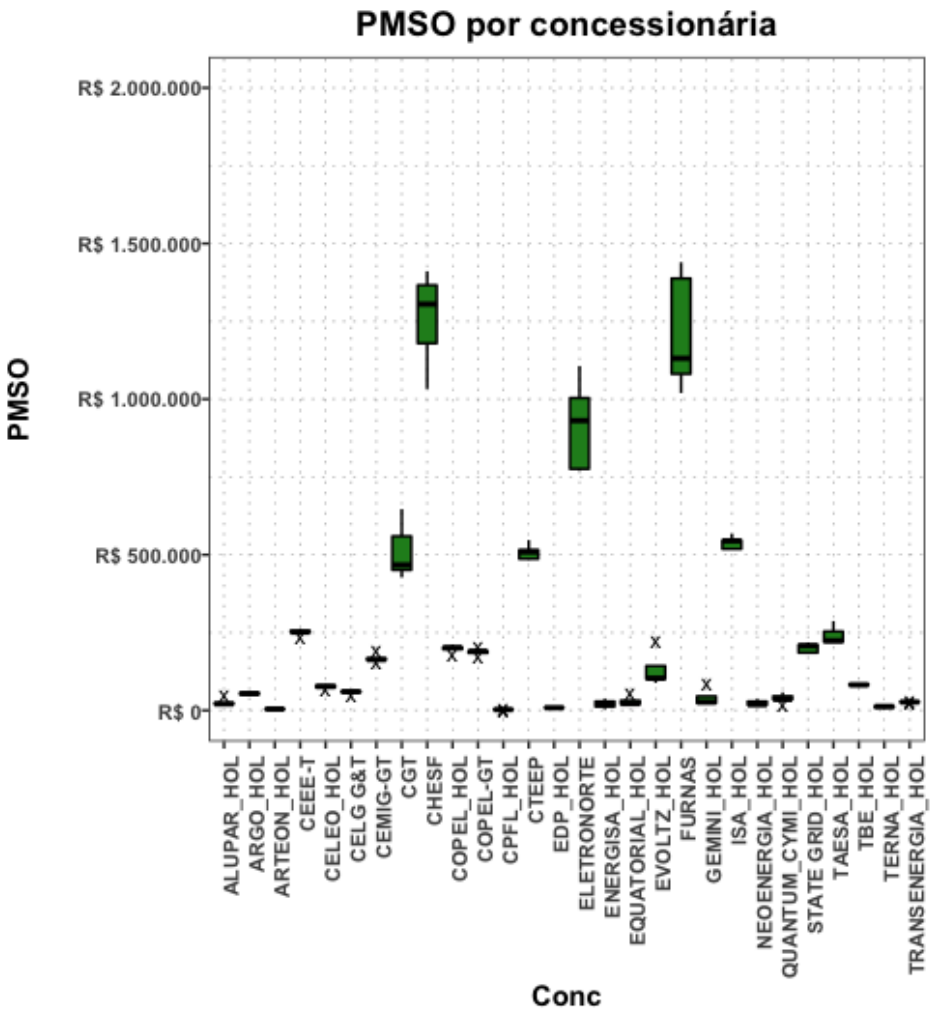


Fonte: Autor.

DEST UFMG

Boxplot por concessionária

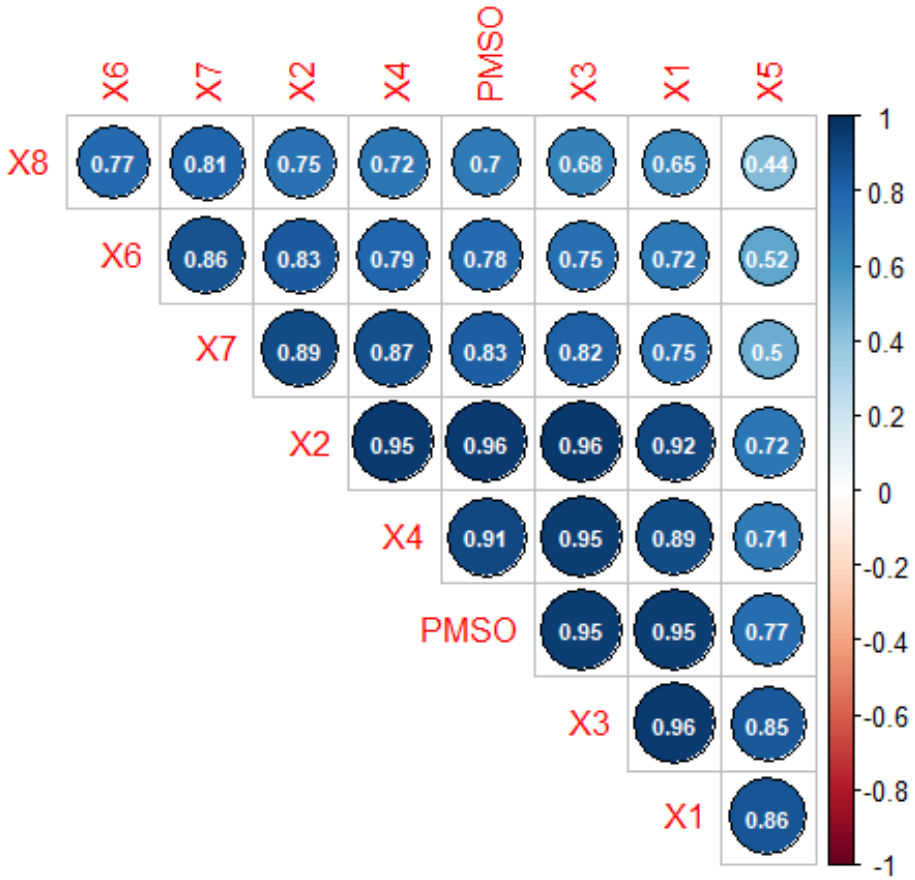
Figura: Boxplot por concessionária



Fonte: Autor.

Correlação

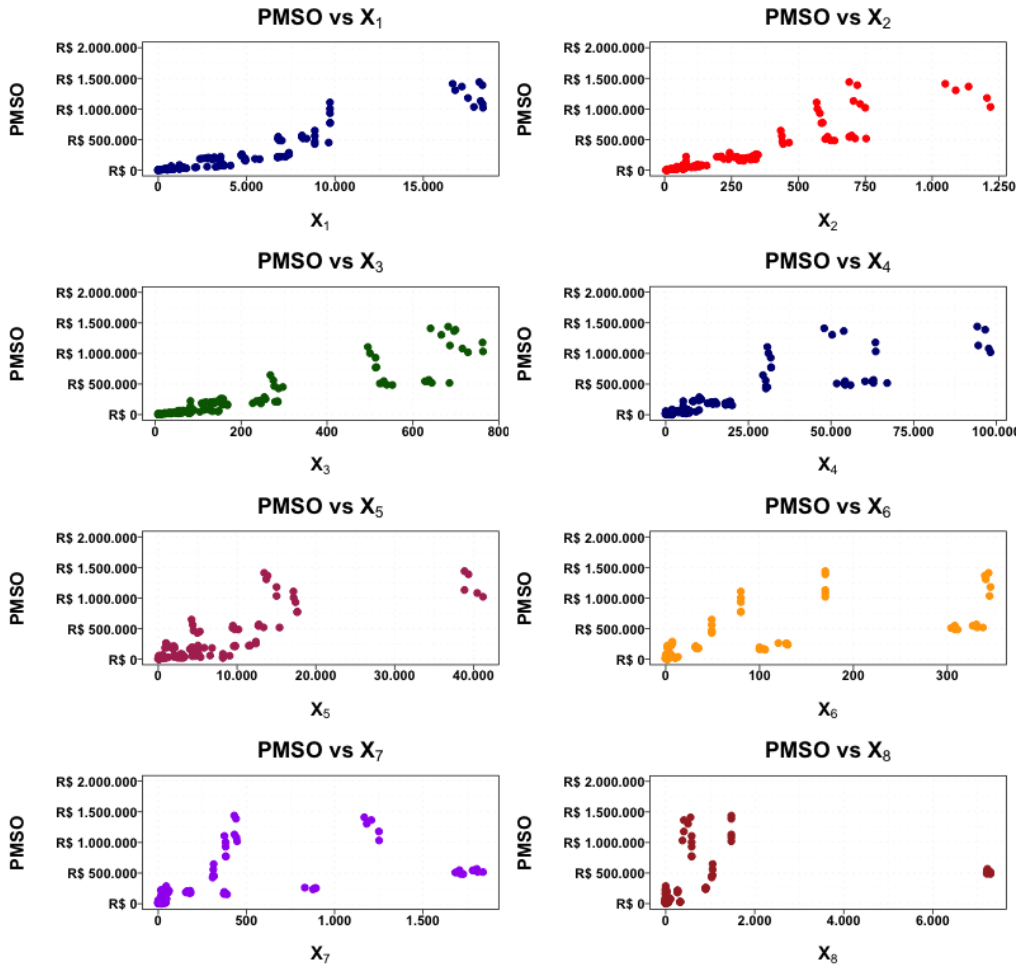
Figura: Matriz de correlação



Fonte: Autor.

Gráficos de dispersão

Figura: Gráficos de dispersão



Fonte: Autor.

Ajuste modelo linear múltiplo

Tabela: Coeficientes do ajuste do modelo

Regressão linear	
	Coeficientes
β_0	-64.039,21
β_1	13,32
β_2	794,1
β_3	169,74
β_4	**-0,52**
β_5	8,86
β_6	621,45
β_7	**-189,72**
β_8	**-8,01**

Fonte: Autor.

Características do ajuste:

- $R^2 = 0.92$
- Modelo ajustado:
  $\hat{y} = -64.03 + 13.32x_1 + 794.1x_2 + 169.7x_3 - 0.5x_4 + 8.8x_5 + 621.4x_6 - 189.7x_7 - 8x_8$

DEST UFMG

Testes dos pressupostos do modelo de regressão

Tabela: Pressupostos do modelo

Testes dos pressupostos do modelo						
Teste	Pressuposto	Estatística	Hipótese nula	valor-p	Significância	**Veredicto**
Teste-F	Significância	$F = 0,92$	$\beta_1 = ...\beta_k = 0$	$2.20 \cdot 10^{-16}$	$0,05$	**Rejeitado**
Breusch-Pagan	Homoscedasticidade	$LM = 2,25$	$\delta_1 = ...\delta_k = 0$	$6,50 \cdot 10^{-8}$	$0,05$	**Rejeitado**
Durbin-Watson	Autocorrelação	$d = 1,07$	Correlação $= 0$	$2,46 \cdot 10^{-10}$	$0,05$	**Rejeitado**

Fonte: Autor.

DEST UFMG

Testes de normalidade para o modelo de regressão

Tabela: Testes de normalidade

Testes de normalidade para os resíduos					
Teste	Estatística	Hipótese nula	valor-p	Significância	**Veredicto**
Shapiro-Wilk	$W = 0,92$	$H_0 : X \sim N$	$5.54 \cdot 10^{-6}$	$0,05$	**Rejeitado**
Anderon-Darling	$A = 2,25$	$H_0 : X \sim N$	$9,76 \cdot 10^{-6}$	$0,05$	**Rejeitado**
Kolmogorov-Smirnov	$D = 0,15$	$H_0 : X \sim N$	$2,03 \cdot 10^{-7}$	$0,05$	**Rejeitado**

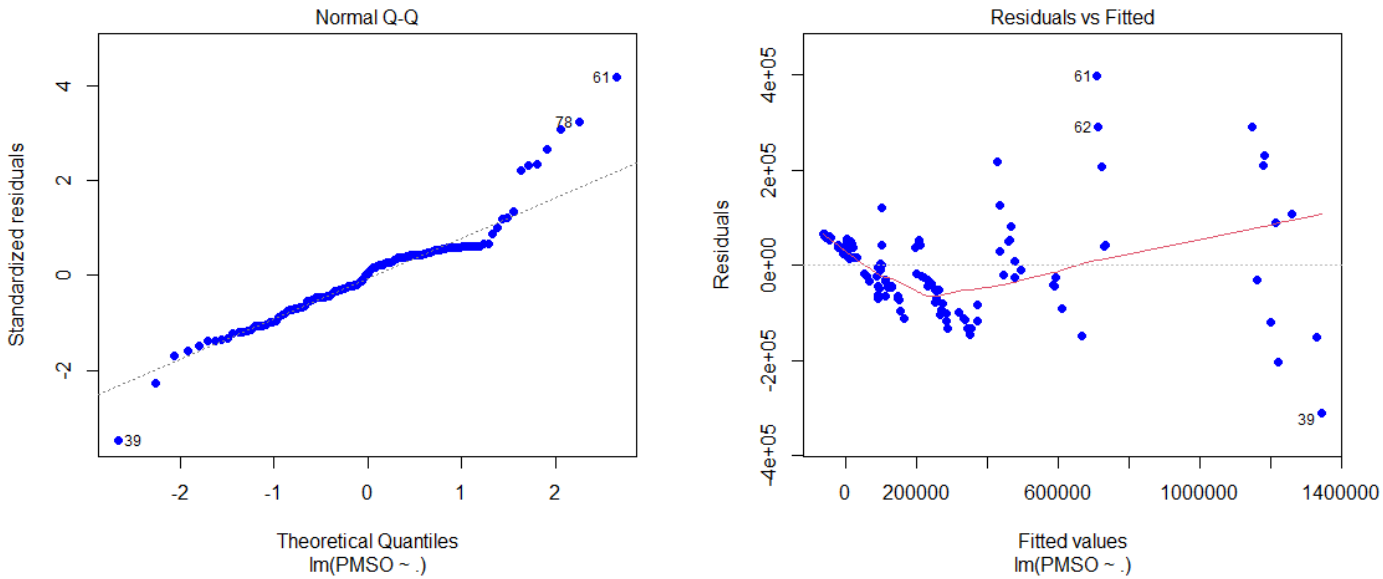
Fonte: Autor.

DEST UFMG

Análise de resíduos

Gráfico de probabilidades $Q - Q$ e gráfico de resíduos *versus* ajuste

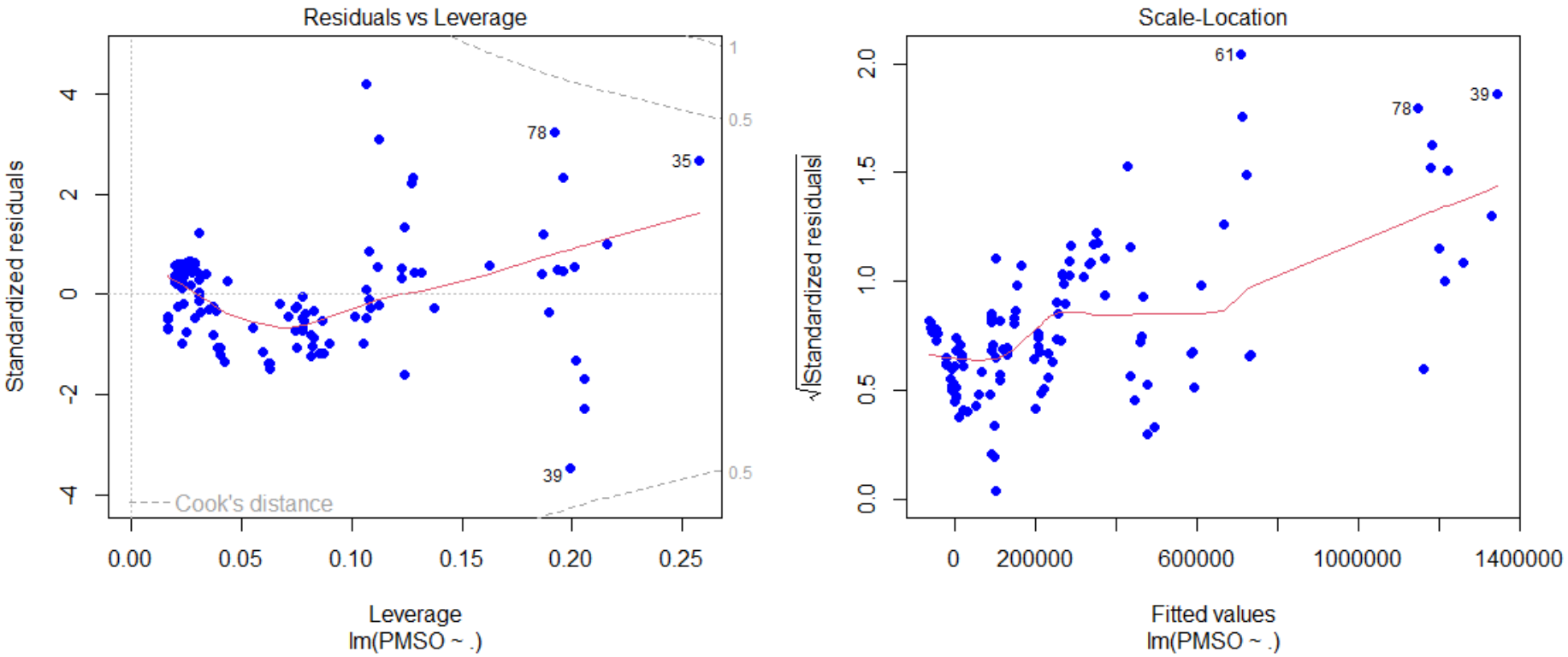
Figura: Gráficos de resíduos



Fonte: Autor.

Análise de resíduos

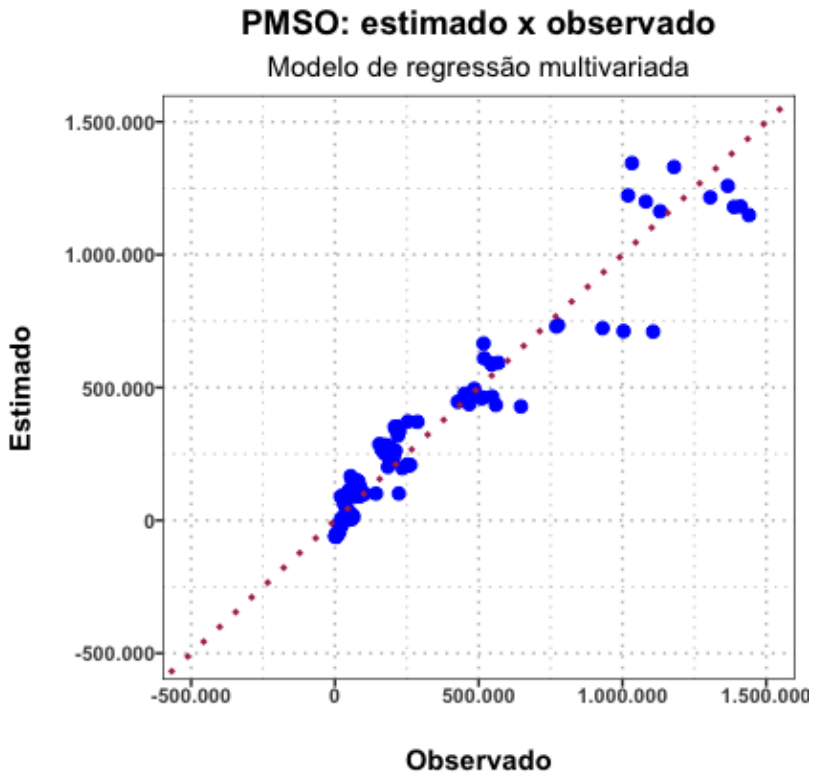
Figura: Gráficos de resíduos



Fonte: Autor.

Ajustes para modelo de regressão multivariado

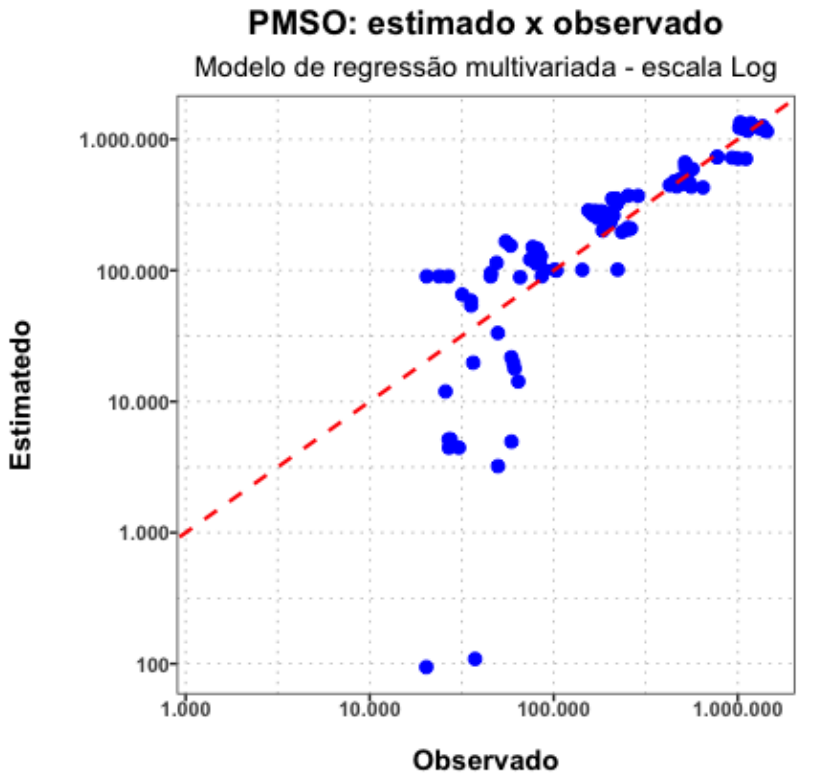
Figura: Modelo multivariado



Fonte: Autor.

R^2 : 0.92

Figura: Modelo multivariado - Log



Fonte: Autor.

R^2 : 0.97

DEST UFMG

Ajuste linear múltiplo leave-one-out com validação cruzada

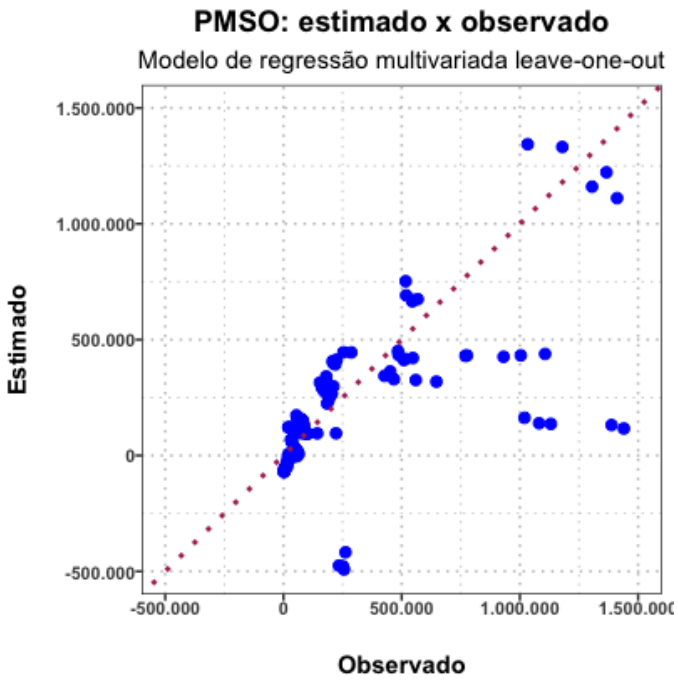
Implementação de validação cruzadas em conjunto com a técnica leave-one-out:

1. Criar um novo banco de dados com informações de uma empresa escolhida: base de dados de validação
2. Criar outro banco de dados com informações das empresas remanescentes: base de treinamento
3. Implementar modelo de regressão linear múltiplo utilizando o banco de dados de treino
4. Implementar a técnica de *AIC* para escolher o melhor modelo e conjunto de variáveis
5. Aplicar o modelo de regressão linear múltiplo na base de dados de validação para estimar os valores de PMSO do modelo
6. Armazenar os dados do PMSO estimado para a empresa do banco de dados de validação
7. Realizar esse procedimento para todas as 28 empresas

DEST UFMG

Ajustes para modelo de regressão *leave-one-out*

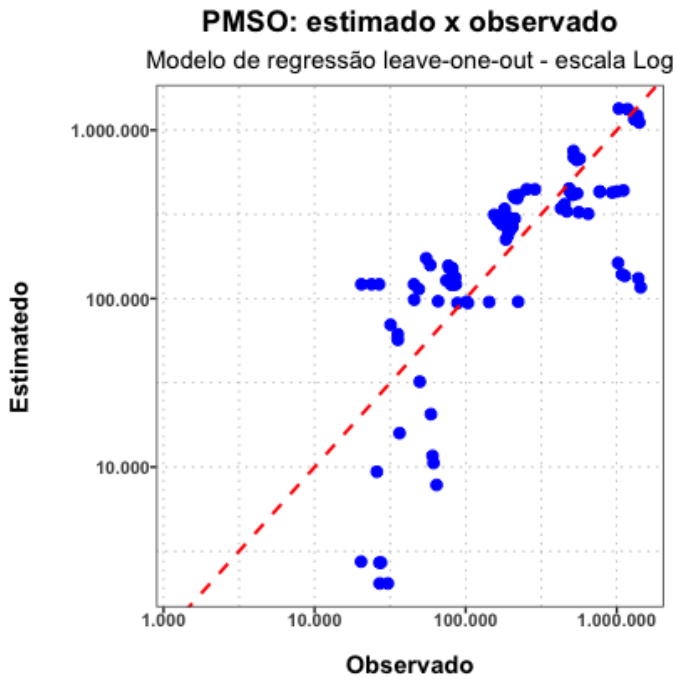
Figura: Modelo multivariado



Fonte: Autor.

R^2 : 0.32

Figura: Modelo multivariado - Log



Fonte: Autor.

R^2 : 0.93

DEST UFMG

Ajuste - modelo gama

Implementação do modelo gama para lidar com os problemas encontrados anteriormente.

Foi escolhido o modelo gama devido a evidência de proporcionalidade quadrática na variância dos resíduos.

Implementação do modelo com as seguintes abordagens:

1. Ajuste com implementação do critério de informação de Akaike (AIC) – Regressão gama (1)

2. Ajuste com implementação do AIC e partição da base de dados em dois, um conjunto para treino o modelo e outro para aplicação do modelo preditivo (análogo ao realizado na seção anterior) – Regressão gama (2)

DEST UFMG

Ajuste - Modelo Gama

Tabela: Ajuste - modelo gama

Resultados - ajuste modelo Gama		
Coeficientes	(1) Step AIC	(2) Step AIC - Partição
β_0	-1.754,9	-1.273,8
β_1	19,6	17,9
β_2	376,0	431,0
β_3	-	-
β_4	-	-
β_5	-	-
β_6	1.006,3	585,7
β_7	-	-
β_8	-25,5	-

Fonte: Autor.

Tabela: Qualidade do ajuste

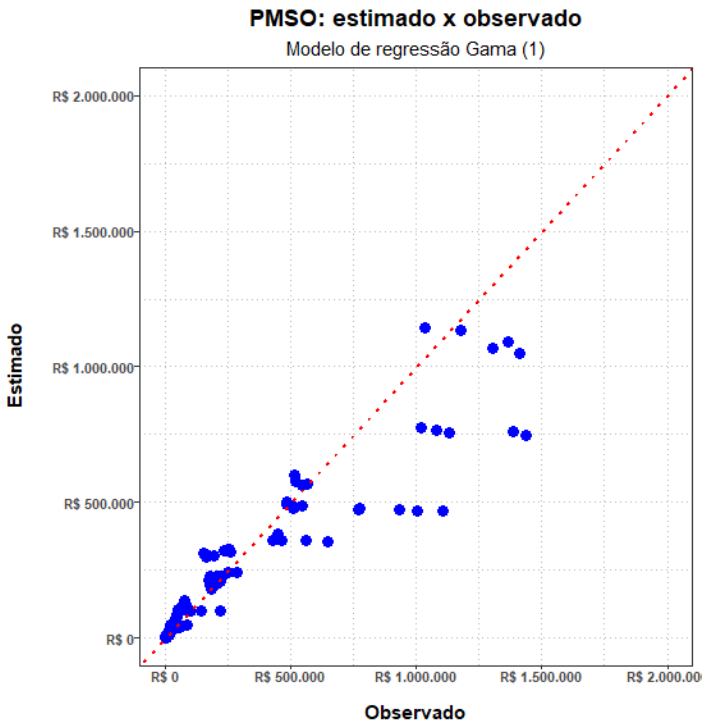
Modelo	R^2	**Deviance**
Regressão Gama (1)	0,82	1,00
Regressão Gama (2)	0,72	1,00

Fonte: Autor.

DEST UFMG

Resultados para modelo de regressão gama

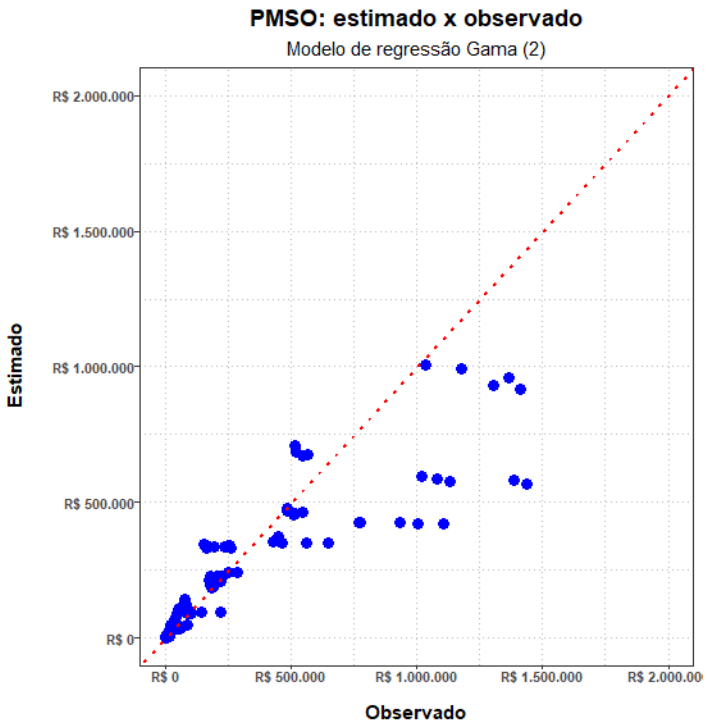
Figura: Modelo (1)



Fonte: Autor.

R^2 : 0.82
p-valor: 1,00

Figura: Modelo (2)



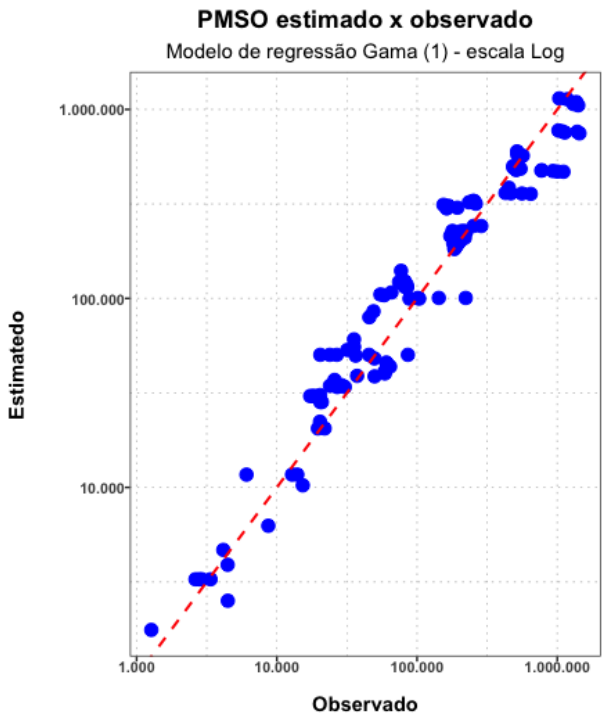
Fonte: Autor.

R^2 : 0.72
p-valor: 1,00

DEST UFMG

Resultados para modelo de regressão gama - logaritmo

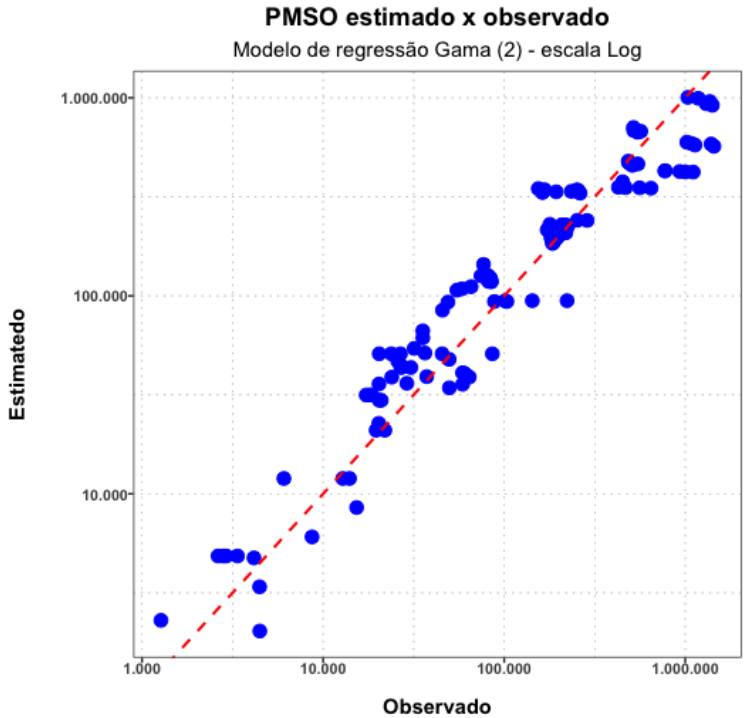
Figura: Modelo (1)



Fonte: Autor.

R^2 : 0.94

Figura: Modelo (2)



Fonte: Autor.

R^2 : 0.92

DEST UFMG

Abaixo, algumas ponderações para implementação do modelo de programação linear:

- Modelo oriundo da regressão quantílica a $y = \beta_0 + \beta x + \varepsilon$ com minimização de erros
- Função objetivo: minizar erros com fator τ
- Restrições:
  - $\beta \geq 0$
  - $\beta_0 = \alpha$
- Particionamento do erro em dois termos
- Implementação de modelo clássico e abordagem *leave-one-out*

DEST UFMG

Formulação do modelo de programação linear

A abordagem escolhida foi a da mediana, portanto $\tau = 0.5$

$$\text{minimizar:} \quad \sum_{j=1}^{n} \tau e_{1j} \; + \; (1-\tau) e_{2j}$$

$$\text{sujeito a:} \quad y_j = \alpha_j + \beta_{kj} x_{kj} + e_{1j} - e_{2j}, \quad \forall j \in \{1, \ldots, n\}$$

$$\beta_k \geq 0, \quad \forall k \in \{1, \ldots, 8\}$$

$$e_j = e_{1j} + e_{2j}, \quad \forall j \in \{1, \ldots, n\}$$

$$N \in \{0, \ldots, 128\}$$

DEST UFMG

Algoritmo *leave-one-out*

1. Remover as observações para a concessionária na qual queremos estimar o PMSO.
2. Resolver o modelo de programação linear utilizando o conjunto de dados com as concessionárias que restaram.
3. Com o output do modelo contendo os coeficientes de interesse, estimar o PMSO para a concessionária de interesse.
4. Armazenar resultado.
5. Realizar o procedimento 28 vezes, número de concessionárias presentes no estudo.

DEST UFMG

Solução modelo linear

Tabela: Solução - programação linear

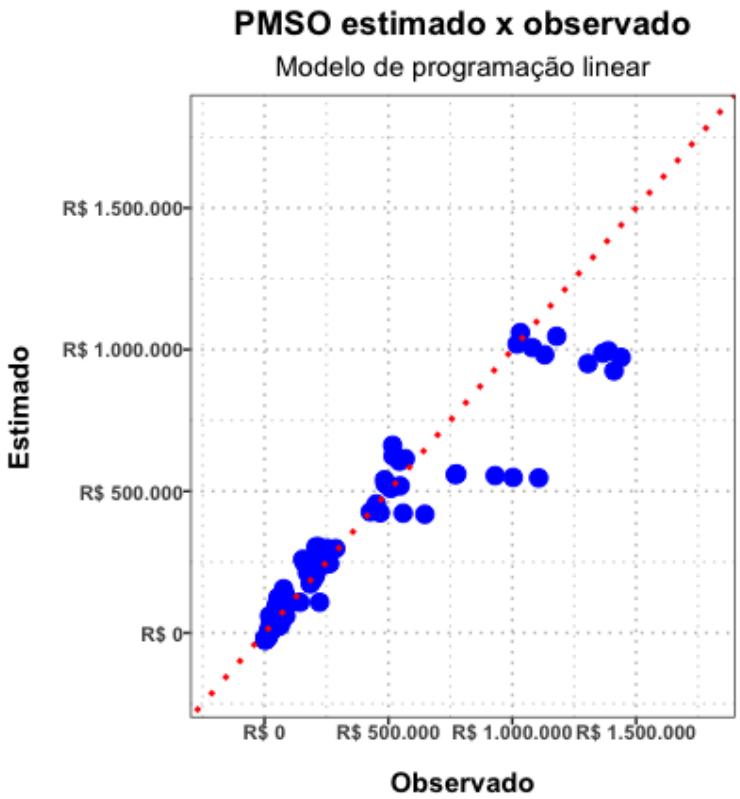
Programação Linear		
Coeficientes	Modelo completo	*Leave-one-out*
β_0	-29.345,2	27.379,5
β_1	20,7	20,9
β_2	450,0	446,3
β_3	-	-
β_4	1,8	1,7
β_5	3,7	3,7
β_6	-	11,9
β_7	-	-
β_8	-	-
R^2	0,88	0,79

Fonte: Autor.

DEST UFMG

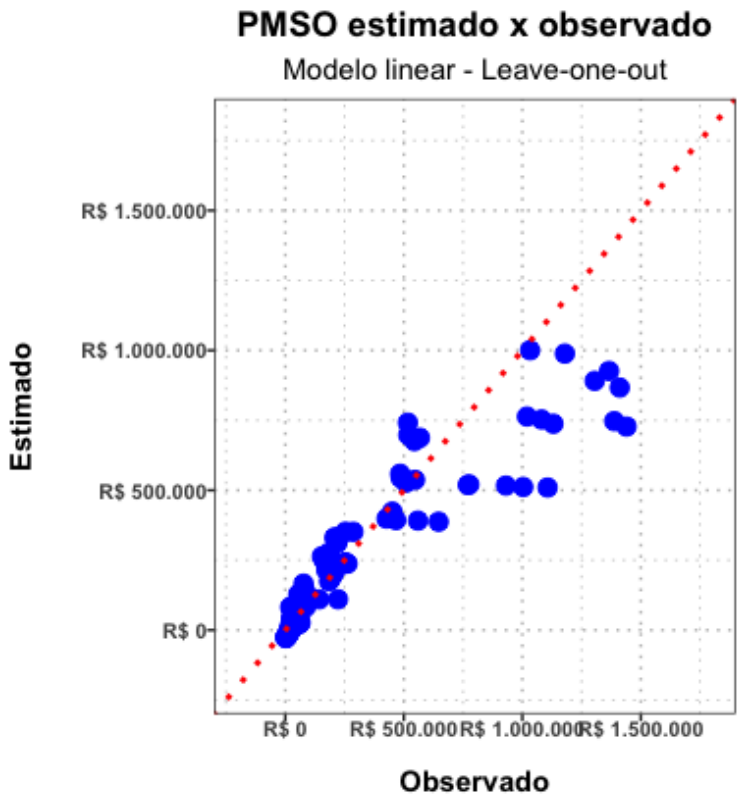
Soluções dos modelos: PMSO estimado *x* observado

Figura: Modelo linear



Fonte: Autor.

Figura: Modelo linear - *leave-one-out*

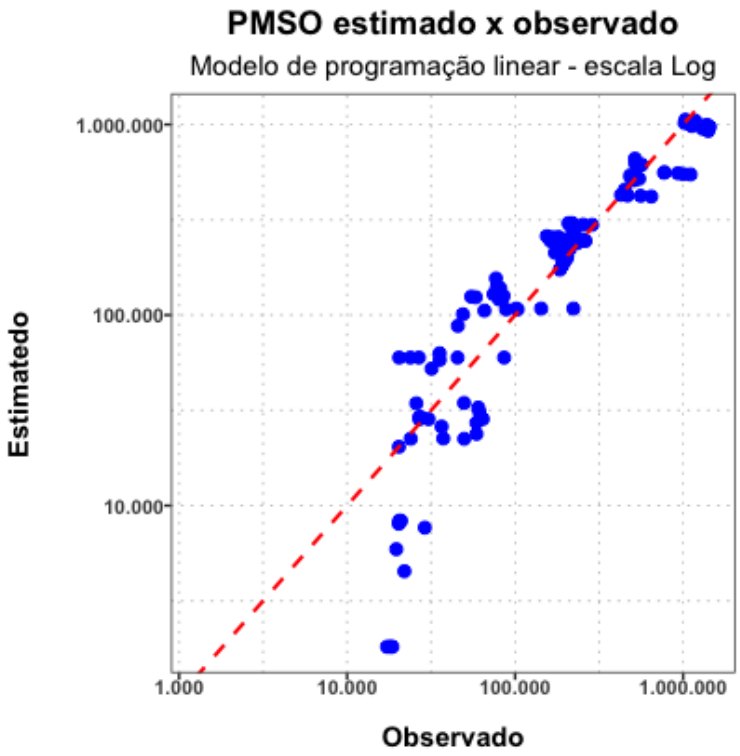


Fonte: Autor.

DEST UFMG

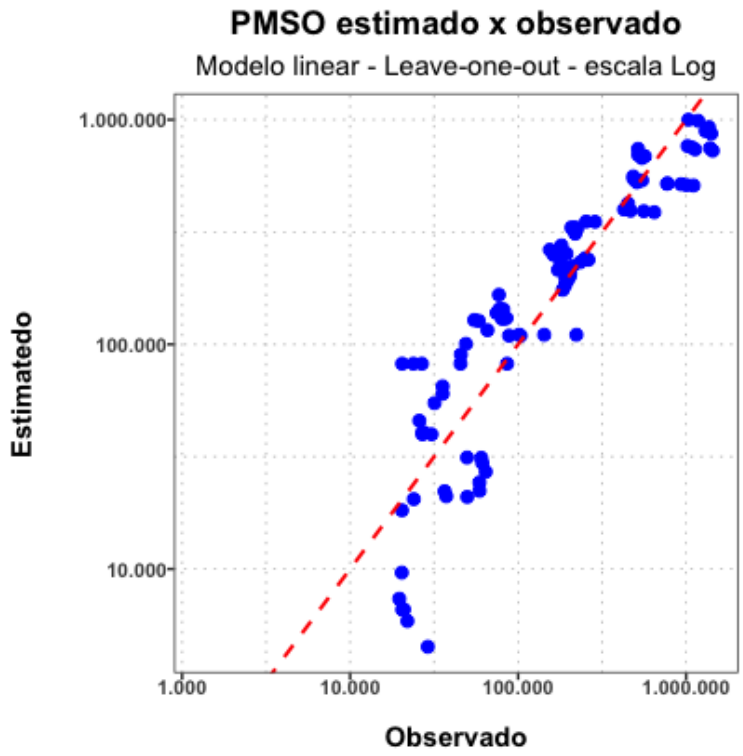
PMSO estimado x observado - logaritmo

Figura: Modelo linear - log



Fonte: Autor.

Figura: *Leave-one-out* - log



Fonte: Autor.

DEST UFMG

Modelo linear - bootstrap

Para a verificar a robustez das restrições do modelo de programação linear foi implementada a técnica de bootstrap e para intervalo de confiança percentílico. O modelo completo foi escolhido por apresentar maior coeficiente de determinação.

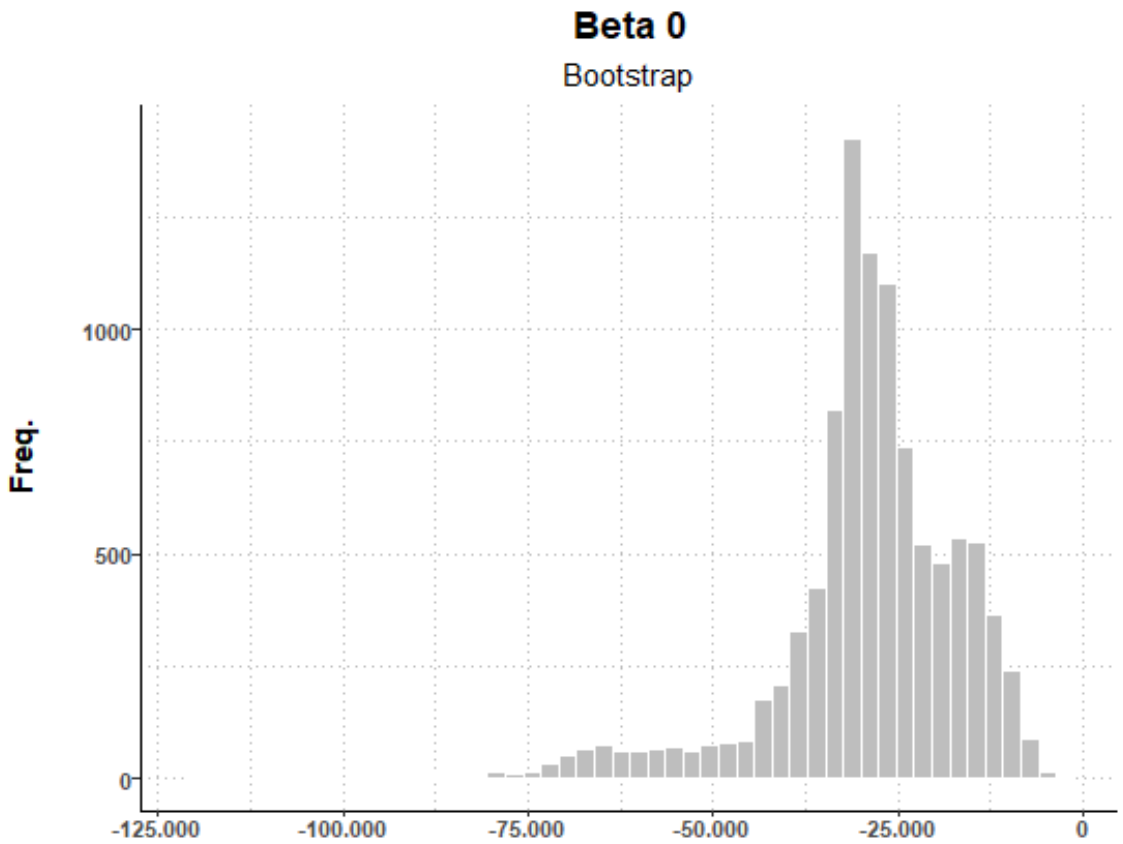
Nesse procedimento foram realizadas 10.000 simulações com observações aleatoriamente no software *R*. Abaixo os passos do procedimento:

1. É gerada uma nova base de dados utilizando a técnica de reamostragem com reposição
2. É implementado o algoritmo SIMPLEX para resolver o modelo com a nova base de dados oriunda da reamostragem
3. A solução contendo os coeficientes de regressão é computada
4. Os passos de 1 a 3 são realizados 10.000 vezes

DEST UFMG

Resultados para procedimento bootstrap - β_0

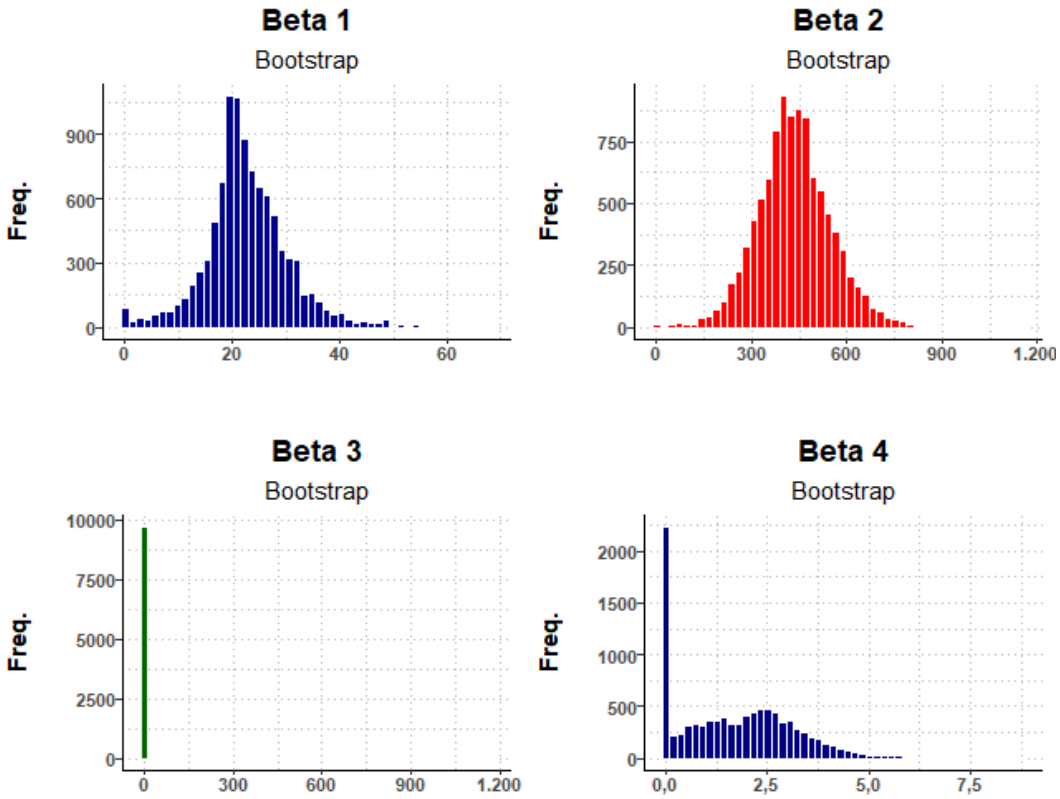
Figura: Histograma de soluções para β_0



Fonte: Autor.

Resultados para procedimento bootstrap - coeficientes

Figura: Histograma de soluções para $\beta_1, \beta_2, \beta_3, \beta_4$

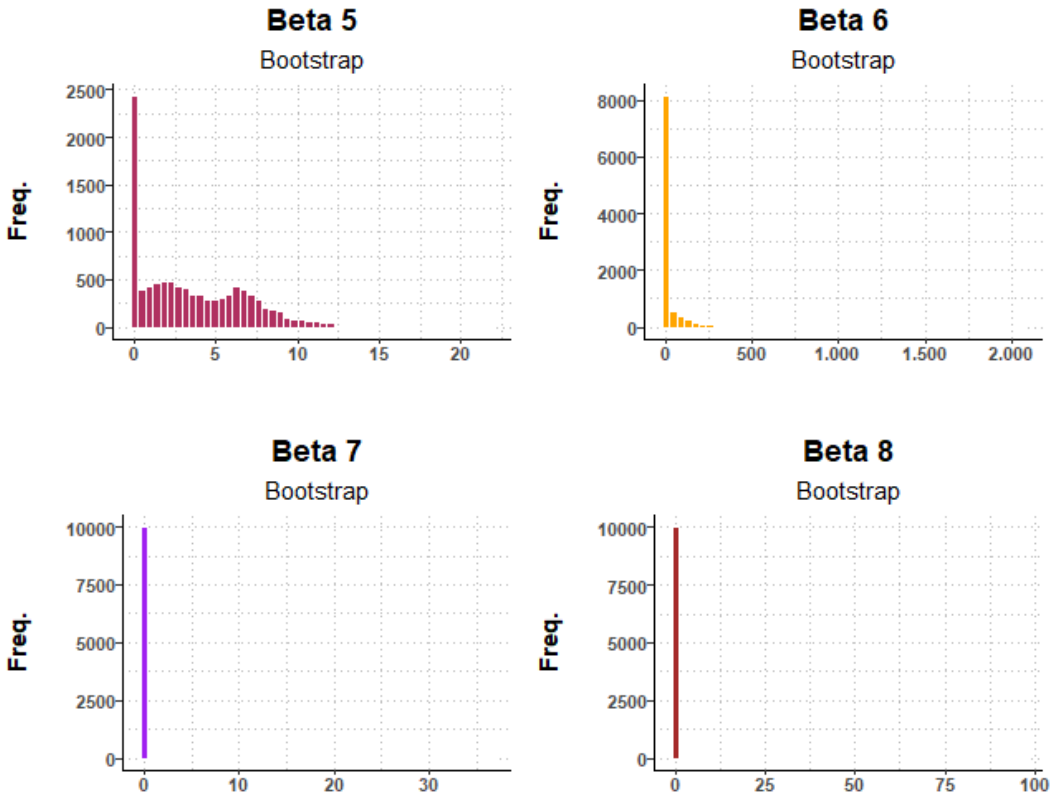


Fonte: Autor.

DEST UFMG

Resultados para procedimento bootstrap - coeficientes

Figura: Histograma de soluções para $\beta_5, \beta_6, \beta_7, \beta_8$



Fonte: Autor.

DEST UFMG

Intervalos de confiança bootstrap

O nível de confiança escolhido para construir o intervalo foi de 95%. Dessa forma, obteve-se os seguintes intervalos com seus limites definidos:

Tabela: Intervalo de Confiança Percentílico - Bootstrap

Intervalo de Confiança Percentílico *Bootstrap*		
Coeficientes	2,5%	97,5%
β_0	-64.426,7	-10.021,2
β_1	6,3	40,0
β_2	206,3	672,2
β_3	0,0	51,3
β_4	0,0	4,6
β_5	0,0	11,1
β_6	0,0	281,0
β_7	0,0	0,0
β_8	0,0	0,0

Fonte: Autor.

DEST UFMG

Resultados agrupados - finais

- Para comparação dos resultados, foram considerados os modelos implementados com as técnicas de validação cruzada e leave-one-out na escala logarítmica.
- Tal abordagem foi utilizada para atenuação das distorções presentes no banco de dados aqui descritas.
- Dessa forma, no estudo, foram considerados os resultados dos ajustes de modelos de **regressão linear**, **regressão não-linear (modelo gama)** e de **programação linear**.

Tabela: Resultados Finais

Comparativo de R^2	
Modelo	R^2
Regressão linear - Leave-one-out - Log	**0,93**
Regressão Gama (2) - Log	**0,92**
Modelo Linear - Leave-one-out - Log	**0,80**

Fonte: Autor.

DEST UFMG

Conclusões

- O ajuste produzido pelo modelo linear múltiplo se mostrou inadequado por violar os pressupostos de normalidade e as restrições operacionais.
- O ajuste produzido pelo modelo gama se mostrou razoável em relação a capacidade preditiva, entretanto apresentou problemas de subdispersão.
- O modelo de programação linear se mostrou **o mais efetivo** para estimar o *PMSO* pois respeita as restrições operacionais e apresenta melhor capacidade preditiva.
- Cinco entre as oito variáveis mostraram-se relevantes (coeficiente posivito e não nulo) quando aplicado o modelo linear.
- **Atualmente a ANEEL implementa uma metodologia que utiliza direcionadores de custo redundantes podendo comprometer ou enviesar as estimativas de eficiência.**

DEST
UFMG