

Exemplo Abastecimento de Refrigerantes

Gilberto A. Paula

Departamento de Estatística
IME-USP, Brasil
giapaula@ime.usp.br

2^o Semestre 2023

- 1 Abastecimento de Refrigerantes
- 2 Análise de Dados Preliminar
- 3 Ajuste Mínimos Quadrados
- 4 Diagnóstico Modelo Ajustado
- 5 Ajuste Modelo Quadrático
- 6 Diagnóstico Modelo Final
- 7 Conclusões
- 8 Referências

Descrição do Experimento

Uma engarrafadora de refrigerantes está analisando o serviço de abastecimento das máquinas de refrigerantes atendidas pela empresa. O serviço de abastecimento inclui o estoque das garrafas nas máquinas e pequenas manutenções feitas pelo próprio motorista do veículo com os carregamentos.

Descrição das Variáveis

O engenheiro industrial responsável pela logística da distribuição dos refrigerantes acredita que o tempo gasto pelo motorista para o abastecimento das máquinas (**tempo**, em minutos) pode estar relacionado com as seguintes variáveis explicativas:

- **distância**, distância percorrida pelo motorista do veículo até as máquinas (em pés^a),
- **ncaixas**, número de caixas de produtos estocados.

Uma amostra aleatória de 25 abastecimentos feitos num *outlet* foi considerada para análise (Montgomery, Peck e Vining, 2012, Cap. 3).

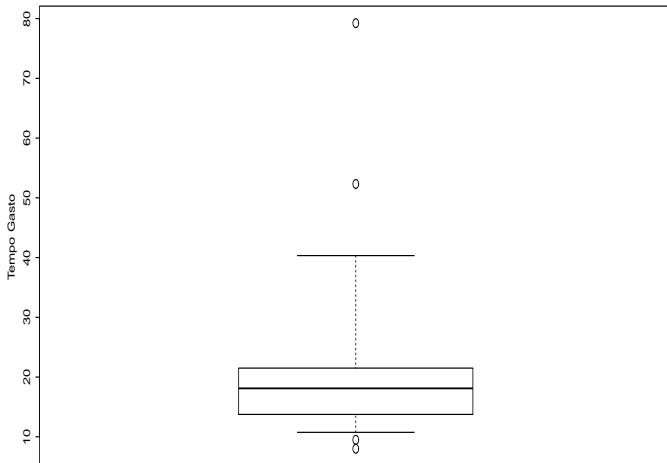
^a1 pé = 0,3048 metros

- 1 Abastecimento de Refrigerantes
- 2 Análise de Dados Preliminar**
- 3 Ajuste Mínimos Quadrados
- 4 Diagnóstico Modelo Ajustado
- 5 Ajuste Modelo Quadrático
- 6 Diagnóstico Modelo Final
- 7 Conclusões
- 8 Referências

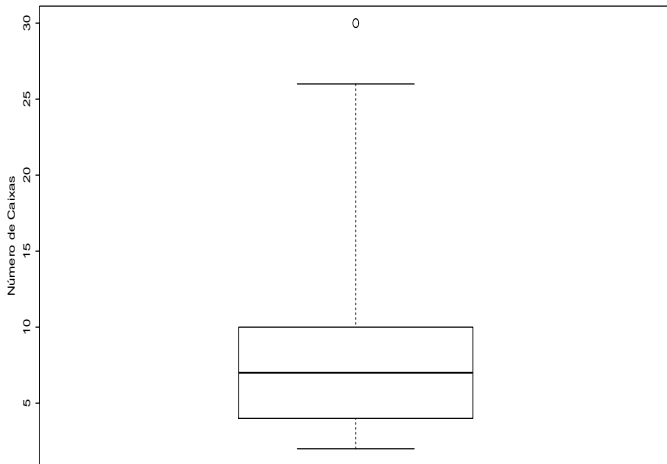
Descrição

Medida	Tempo	Ncaixas	Distância
n	25	25	25
Média	22,38	8,76	409,30
D.Padrão	15,52	6,88	325,19
CV	69%	78%	79%
Mínimo	8,00	2,00	36,00
1 ^o Quartil	13,75	4,00	150,00
Mediana	18,11	7,00	330,00
3 ^o Quartil	21,50	10,00	605,00
Máximo	79,24	30,00	1460,00

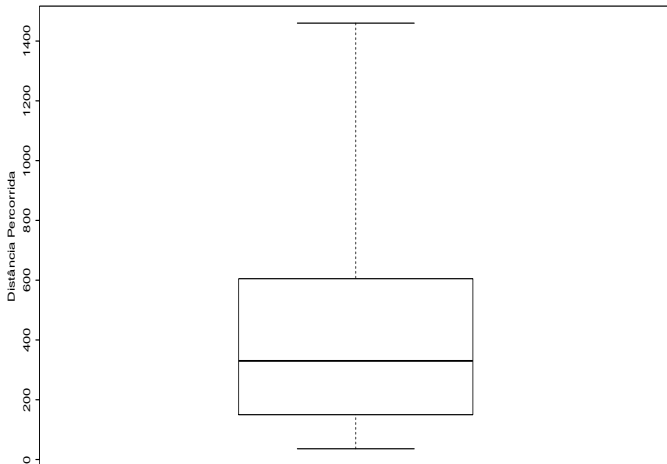
Boxplot Tempo Gasto



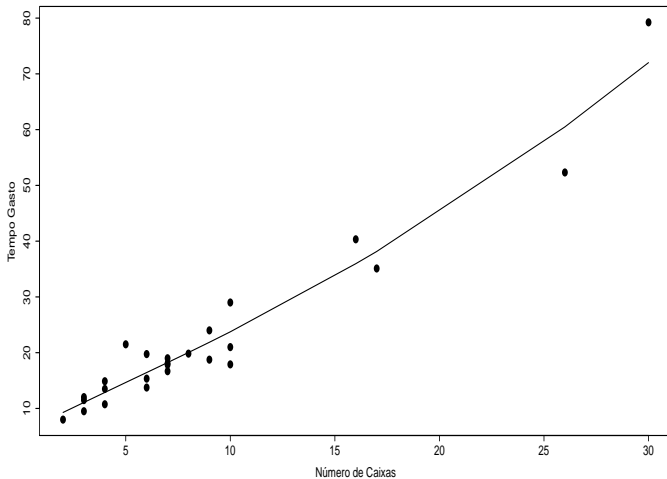
Boxplot Número de Caixas de Produtos



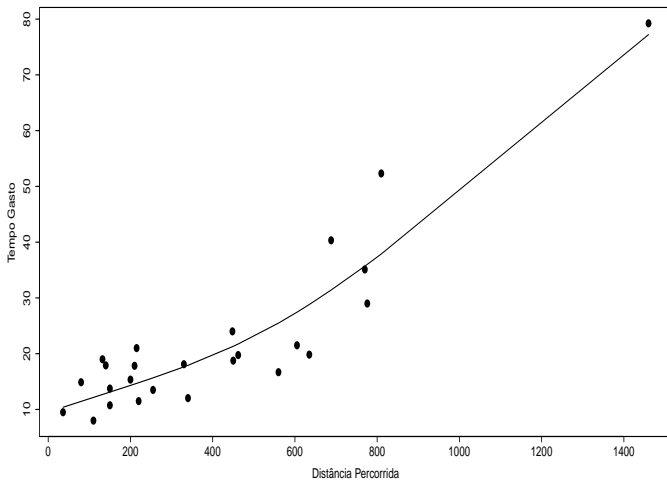
Boxplot Distância Percorrida



Dispersão Tempo Gasto versus Número de Caixas



Dispersão Tempo Gasto versus Distância Percorrida



- 1 Abastecimento de Refrigerantes
- 2 Análise de Dados Preliminar
- 3 Ajuste Mínimos Quadrados**
- 4 Diagnóstico Modelo Ajustado
- 5 Ajuste Modelo Quadrático
- 6 Diagnóstico Modelo Final
- 7 Conclusões
- 8 Referências

Descrição

Nota-se indícios de aumento do tempo gasto pelo motorista com o aumento da distância percorrida e aumento do número de caixas de produtos, sugerindo inicialmente o seguinte modelo linear:

$$y_i = \beta_1 + \beta_2 \times \text{ncaixas}_i + \beta_3 \times \text{distância}_i + \epsilon_i,$$

para $i = 1, \dots, 25$, em que y_i denota o tempo gasto pelo i -ésimo motorista com $\epsilon_j \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.

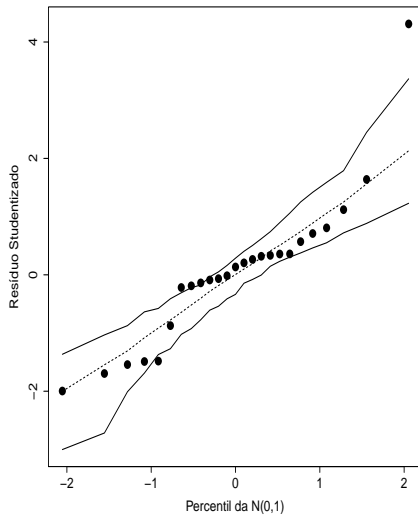
Descrição

As estimativas dos parâmetros são dadas abaixo.

Efeito	Estimativa	Erro padrão	valor-t	valor-P
Constante	2,341	1,097	2,13	0,044
Ncaixas	1,616	0,171	9,47	0,001
Distância	0,014	0,004	3,89	0,000
s	3,259			
R ²	0,96			
\bar{R}^2	0,96			

Todos os parâmetros são marginalmente significativos ao nível de 5%.

- 1 Abastecimento de Refrigerantes
- 2 Análise de Dados Preliminar
- 3 Ajuste Mínimos Quadrados
- 4 Diagnóstico Modelo Ajustado**
- 5 Ajuste Modelo Quadrático
- 6 Diagnóstico Modelo Final
- 7 Conclusões
- 8 Referências



Resíduos Modelo Ajustado

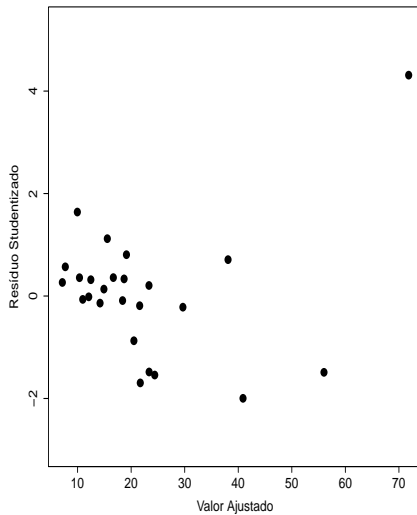


Gráfico da Variável Adicionada Ncaixas

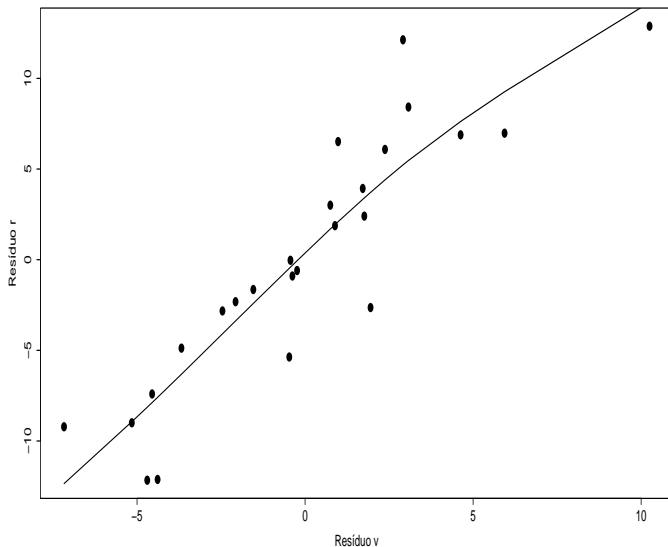
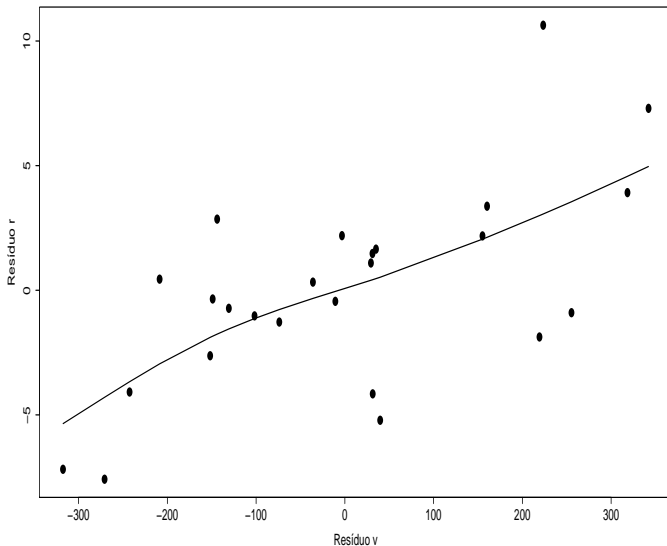
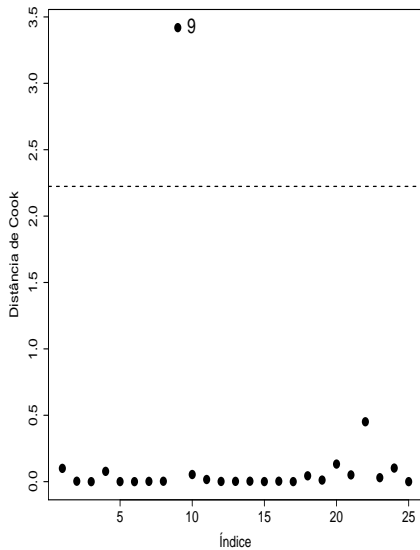


Gráfico da Variável Adicionada Distância



Qualidade do Ajuste

- Pelo gráfico normal de probabilidades nota-se uma tendência sistemática dentro da banda de confiança e um ponto aberrante.
- Nota-se pelo segundo gráfico de resíduos que a variabilidade não está totalmente controlada.
- Há indícios de que pelo menos a variável distância percorrida poderia entrar de forma não linear.



Descrição

Estimativas dos parâmetros eliminando a observação #9.

Efeito	Estimativa	Erro padrão	valor-t	valor-P
Constante	4,447	0,953	4,67	0,000
Ncaixas	1,498	0,130	11,51	0,001
Distância	0,010	0,003	3,57	0,000
s	2,43			
R ²	0,95			
\bar{R}^2	0,94			

Todos os parâmetros são marginalmente significativos ao nível de 1%.

Comentário

- A observação #9 aparece como influente e aberrante. Eliminação dessa observação causa variações desproporcionais nas três estimativas e aumenta a significância da constante. Os coeficientes estimados apresentam variações de $((1,498 - 1,616)/1,616) \times 100\% = -7,3\%$ e $((0,010 - 0,014)/0,014) \times 100\% = -28,6\%$, respectivamente.
- A observação #9 refere-se ao motorista com a maior distância percorrida, maior tempo gasto e maior número de caixas estocadas.

- 1 Abastecimento de Refrigerantes
- 2 Análise de Dados Preliminar
- 3 Ajuste Mínimos Quadrados
- 4 Diagnóstico Modelo Ajustado
- 5 Ajuste Modelo Quadrático**
- 6 Diagnóstico Modelo Final
- 7 Conclusões
- 8 Referências

Inclusão Efeito Quadrático

A fim de tentar acomodar a observação discrepante vamos incluir termo quadrático da distância percorrida conforme sugerido na análise descritiva:

$$y_i = \beta_1 + \beta_2 \times \text{ncaixas}_i + \beta_3 \times \text{ndistância}_i + \beta_4 \times \text{ndistância}_i^2 + \epsilon_i,$$

para $i = 1, \dots, 25$, em que y_i denota o tempo gasto pelo i -ésimo motorista com $\epsilon_j \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ e $\text{ndistância} = \text{distância}/1000$.

Inclusão Efeito Quadrático

Estimativas incluindo efeito quadrático da distância percorrida.

Efeito	Estimativa	Erro padrão	valor-t	valor-P
Constante	6,343	1,211	5,24	0,000
Ncaixas	1,416	0,134	10,61	0,000
nDistância	-0,798	4,336	-0,18	0,856
nDistância ²	14,730	3,324	4,43	0,000
s	2,398			
R ²	0,98			
\bar{R}^2	0,98			

Todos efeitos altamente significativos, exceto efeito linear de nDistância.

Inclusão Efeito Quadrático

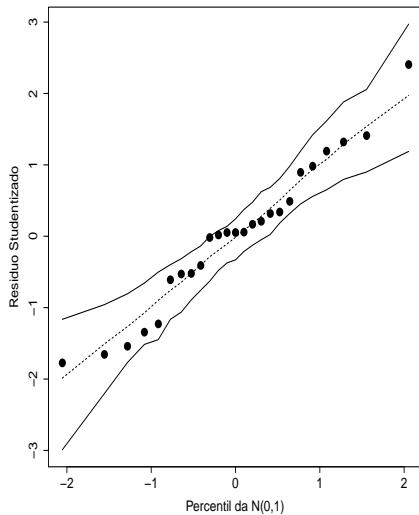
Estimativas modelo final com efeito quadrático em nDistância.

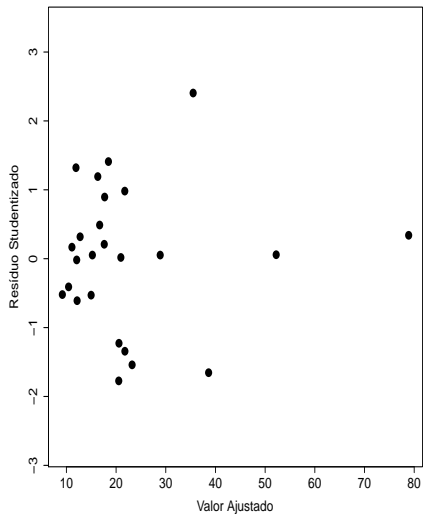
Efeito	Estimativa	Erro padrão	valor-t	valor-P
Constante	6,191	0,867	7,14	0,000
Ncaixas	1,411	0,128	11,05	0,000
nDistância ²	14,247	1,992	7,11	0,000
s	2,345			
R ²	0,98			
\bar{R}^2	0,98			

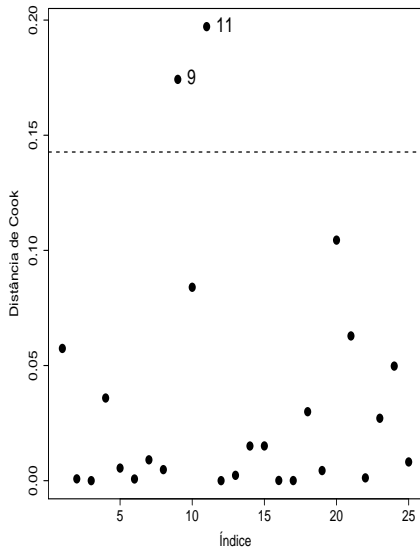
Todos efeitos altamente significativos.

- 1 Abastecimento de Refrigerantes
- 2 Análise de Dados Preliminar
- 3 Ajuste Mínimos Quadrados
- 4 Diagnóstico Modelo Ajustado
- 5 Ajuste Modelo Quadrático
- 6 Diagnóstico Modelo Final**
- 7 Conclusões
- 8 Referências

Resíduos Modelo Ajustado



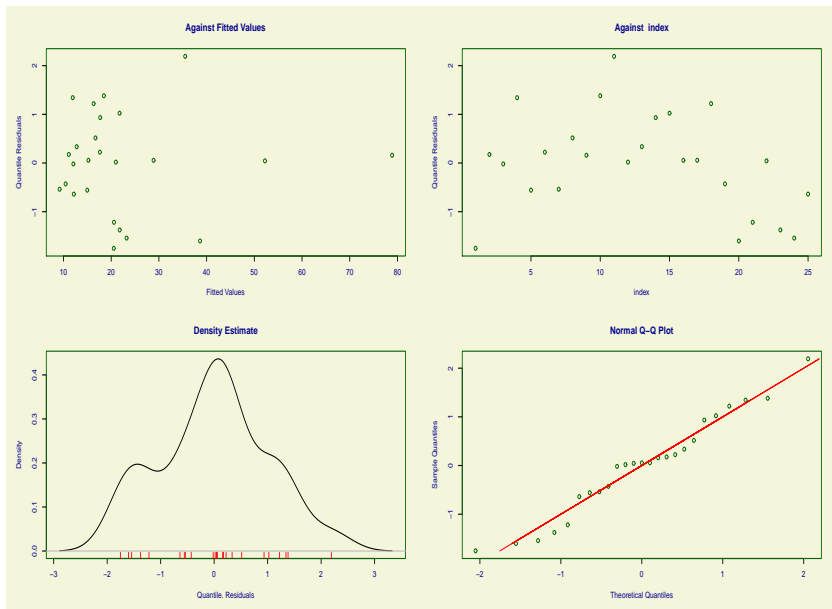




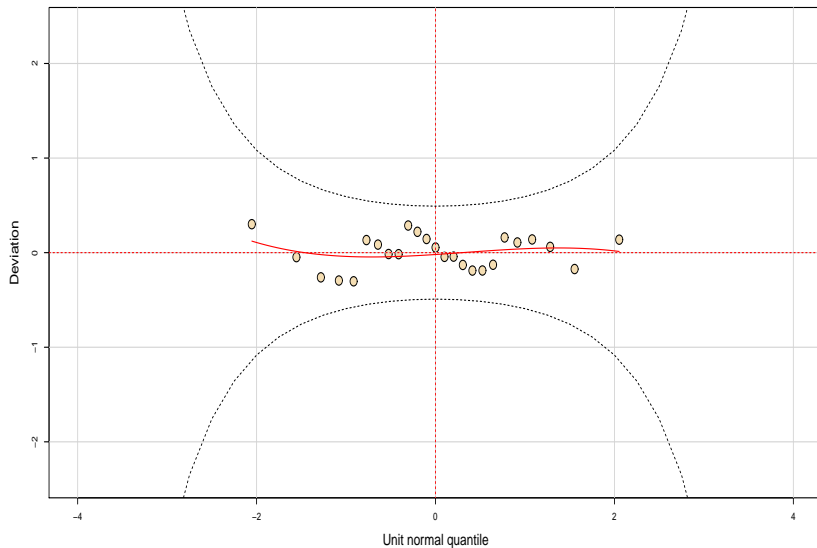
Qualidade do Ajuste

- Pelo gráfico normal de probabilidades nota-se todos os pontos dentro da banda de confiança de forma aleatória. Pelo segundo gráfico de resíduos houve uma melhora no controle da variabilidade.
- A distância de Cook destaca as observações #9 e #11.
- O impacto da eliminação da observação #9 nas estimativas dos coeficientes é de $((1,428 - 1,411)/1,411) \times 100\% = 1,20\%$ e $((13,267 - 14,247)/14,247) \times 100\% = -6,88\%$, respectivamente.
- O impacto da eliminação da observação #11 nas estimativas dos coeficientes é de $((1,338 - 1,411)/1,411) \times 100\% = -5,17\%$ e $((14,969 - 14,247)/14,247) \times 100\% = 5,07\%$, respectivamente.
- Não houve mudanças inferenciais, os coeficientes continuam significativos.

Saída Resíduos pelo GAMLSS



Saída Worm Plot pelo GAMLSS



Interpretação

- Para cada aumento de 1 caixa para estocar, espera-se aumento de 1,411 minutos no tempo gasto pelo motorista (mantendo-se fixa a distância percorrida).
- Porém, o tempo esperado pelo motorista aumenta à medida que aumenta de forma quadrática a distância percorrida (mantendo-se fixo o número de caixas).

- 1 Abastecimento de Refrigerantes
- 2 Análise de Dados Preliminar
- 3 Ajuste Mínimos Quadrados
- 4 Diagnóstico Modelo Ajustado
- 5 Ajuste Modelo Quadrático
- 6 Diagnóstico Modelo Final
- 7 Conclusões**
- 8 Referências

Considerações Finais

- Este é um exemplo de regressão linear múltipla que sob o ajuste de mínimos quadrados com efeitos lineares há presença de uma observação discrepante com maior destaque e cuja eliminação causa variações desproporcionais nas estimativas.
- Porém, levando-se em conta um termo quadrático para a distância percorrida o modelo fica bem ajustado.
- As saídas de resíduos pelo GAMLSS confirmam a qualidade do ajuste.

- 1 Abastecimento de Refrigerantes
- 2 Análise de Dados Preliminar
- 3 Ajuste Mínimos Quadrados
- 4 Diagnóstico Modelo Ajustado
- 5 Ajuste Modelo Quadrático
- 6 Diagnóstico Modelo Final
- 7 Conclusões
- 8 Referências**

Referência

- Montgomery, D. C.; Peck, E. A. e Vining, G. G. (2021). *Introduction to Linear Regression Analysis, 6th Edition*. Hoboken: Wiley.
- Weisberg, S. (2014). *Applied Linear Regression, 4th Edition*. Wiley