

Faturamento de Restaurantes

Gilberto A. Paula

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

1^o Semestre 2023

- 1 Faturamento de Restaurantes
- 2 Análise de Dados Preliminar
- 3 Regressão Linear Simples
- 4 Regressão Ponderada
- 5 Ajuste Modelo Duplo
- 6 Conclusões
- 7 Referências

Descrição dos Dados

Considere uma amostra aleatória de $n = 30$ restaurantes em que observou-se para cada um as seguintes variáveis aleatórias:

- **faturamento anual (em mil USD)**
- **gasto anual com publicidade (em mil USD).**

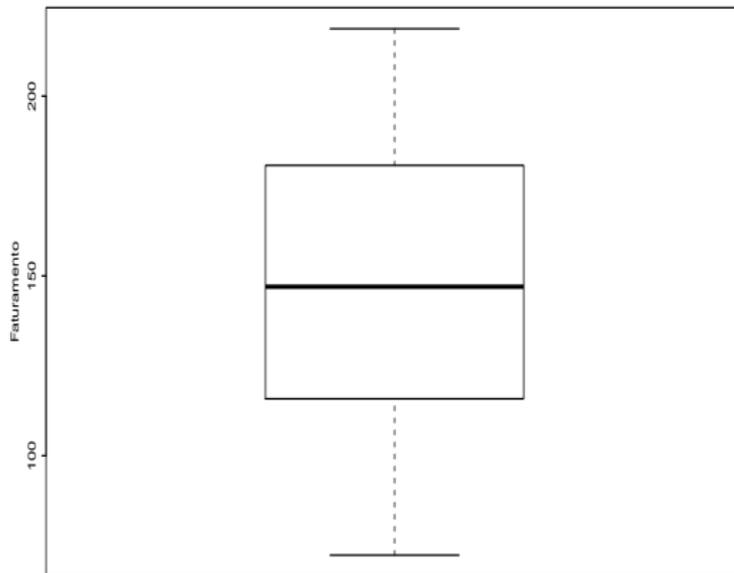
O objetivo principal do estudo é tentar relacionar o faturamento médio anual com o gasto anual com publicidade (Montgomery, Peck e Vining, 2021).

- 1 Faturamento de Restaurantes
- 2 Análise de Dados Preliminar**
- 3 Regressão Linear Simples
- 4 Regressão Ponderada
- 5 Ajuste Modelo Duplo
- 6 Conclusões
- 7 Referências

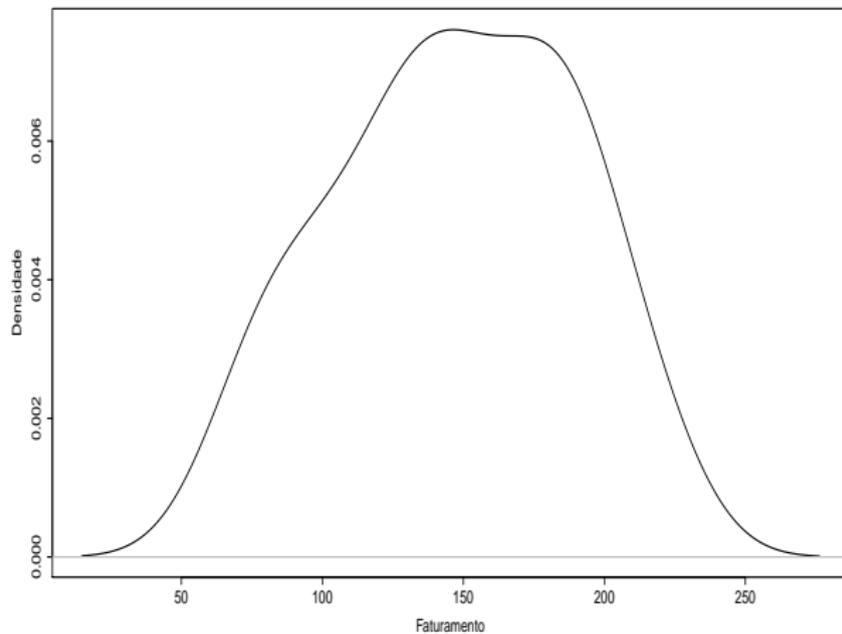
Descrição

Medida	Faturamento	G. Publicidade
n	30	30
Média	147,10	12,13
D.Padrão	42,13	5,12
CV	28,6%	42,2%
Mínimo	72,34	3,00
1 ^o Quartil	117,70	8,93
Mediana	147,00	12,46
3 ^o Quartil	180,30	15,19
Máximo	218,70	19,50

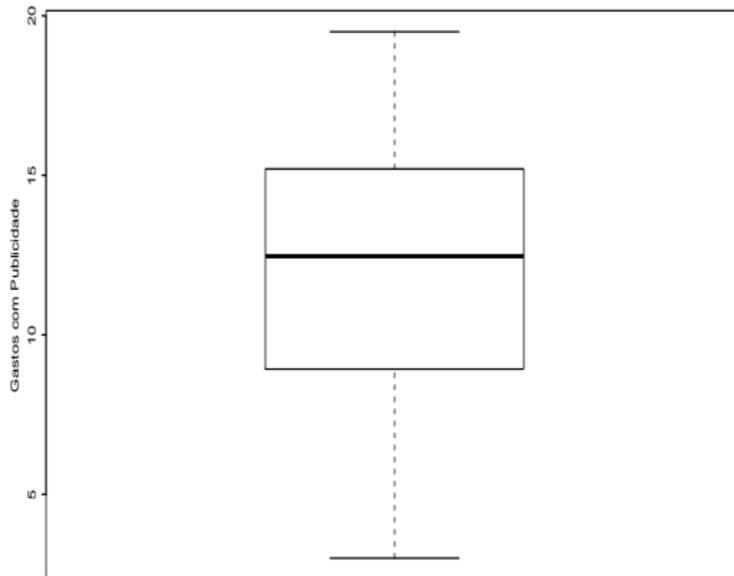
Boxplot Faturamento Anual



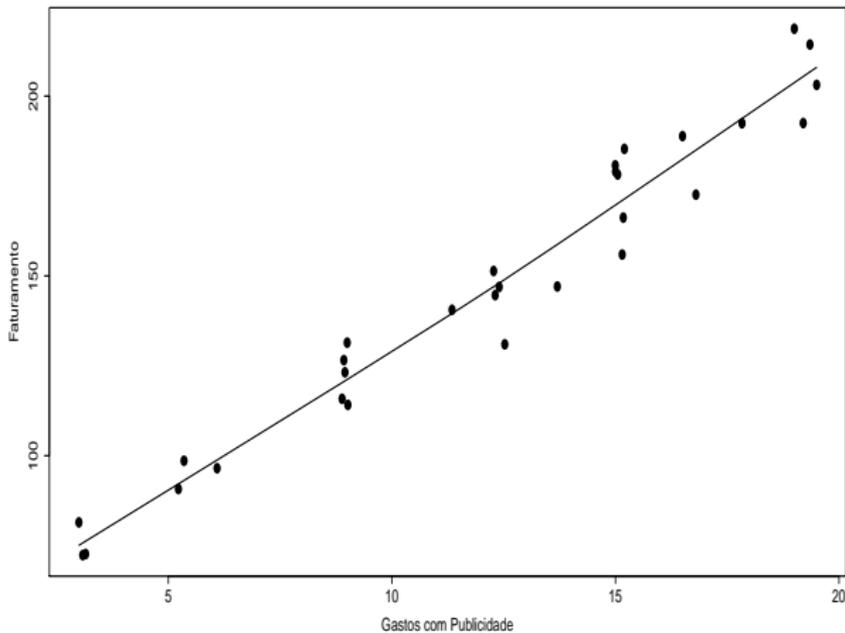
Densidade Faturamento Anual



Boxplot Gastos com Publicidade



Dispersão Faturamento versus Publicidade



- 1 Faturamento de Restaurantes
- 2 Análise de Dados Preliminar
- 3 Regressão Linear Simples**
- 4 Regressão Ponderada
- 5 Ajuste Modelo Duplo
- 6 Conclusões
- 7 Referências

Descrição

Nota-se indícios de aumento do faturamento com o aumento do gasto com publicidade. Essa tendência sugere inicialmente o seguinte modelo de regressão linear:

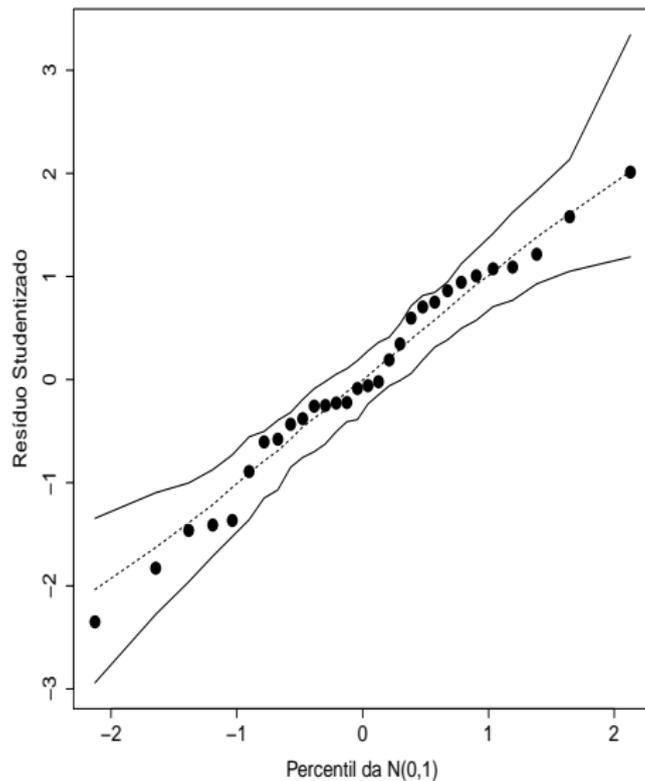
$$y_i = \beta_1 + \beta_2 \times \text{publicidade}_i + \epsilon_i,$$

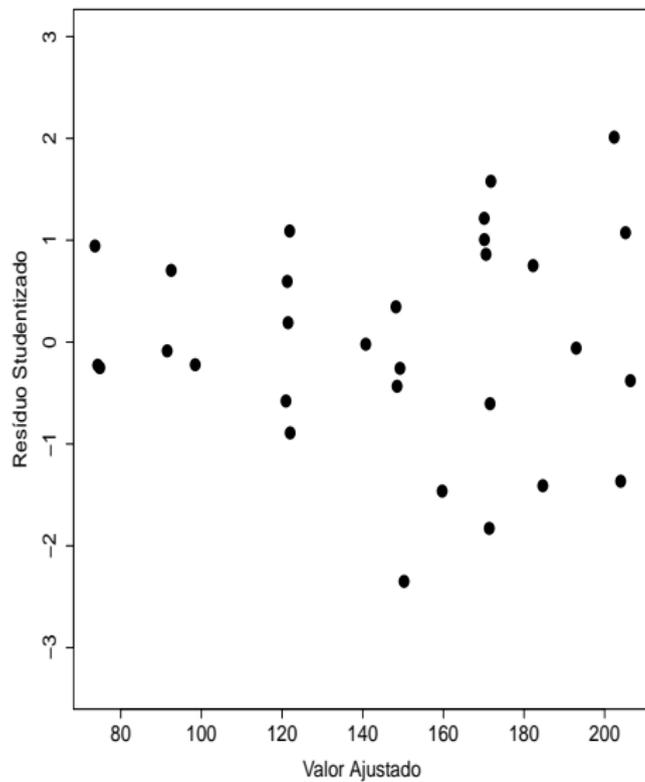
para $i = 1, \dots, 30$, em que y_i denota o faturamento anual do i -ésimo restaurante com $\epsilon_j \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.

Descrição das Estimativas

Efeito	Estimativa	E.Padrão	valor-t	valor-P
Constante	49,443	4,289	11,53	0,00
Publicidade	8,048	0,326	24,68	0,00
s	9,000			
R ²	0,955			
\bar{R}^2	0,954			

Os coeficientes são marginalmente significativos.





Comentários

Embora o ajuste mostra-se adequado através do gráfico normal de probabilidades, nota-se pelo gráfico do resíduo Studentizado contra o valor ajustado indícios de que a variância dos erros não é constante.

Há um aumento no valor do resíduo com o aumento no valor ajustado.

- 1 Faturamento de Restaurantes
- 2 Análise de Dados Preliminar
- 3 Regressão Linear Simples
- 4 Regressão Ponderada**
- 5 Ajuste Modelo Duplo
- 6 Conclusões
- 7 Referências

Descrição

A fim de controlar a variabilidade é proposto o seguinte modelo de regressão ponderada:

$$y_i = \beta_1 + \beta_2 \times \text{publicidade}_i + \epsilon_i,$$

em que y_i denota o faturamento anual do i -ésimo restaurante com $\epsilon_i \stackrel{\text{ind}}{\sim} \mathbf{N}(0, \sigma_i^2)$, $\sigma_i^2 = \sigma^2 \omega_i$ e $\omega_i = \widehat{s}_{y_i}^2$ (Montgomery, Peck e Vining, 2021), para $i = 1, \dots, 30$.

TABLE 5.9 Restaurant Food Sales Data

Notação Flula $w_i = \hat{\beta}_y^2$

Obs. i	(a) Income, Y_i	(b) Advertising Expense, X_i	(c) \bar{x}	(d) s_y^2	(e) Weights, w_i
1	81,464	3,000	3,078.3	26,794,616	6.21771 $E-08$
2	72,661	3,150			5.79507 $E-08$
3	72,344	3,085			5.97094 $E-08$
4	90,743	5,225	5,287.5	30,772,013	2.98667 $E-08$
5	98,588	5,350			2.90195 $E-08$
6	96,507	6,090			2.48471 $E-08$
7	126,574	8,925	8,955.0	52,803,695	1.60217 $E-08$
8	114,133	9,015			1.58431 $E-08$
9	115,814	8,885			1.61024 $E-08$
10	123,181	8,950	12,171.0	59,646,475	1.59717 $E-08$
11	131,434	9,000			1.58726 $E-08$
12	140,564	1,1345			1.22942 $E-08$
13	151,352	12,275	15,095.0	120,571,061	1.12852 $E-08$
14	146,926	12,400			1.11621 $E-08$
15	130,963	12,525			1.10416 $E-08$
16	144,630	12,310	16,650.0	132,388,992	1.12505 $E-08$
17	147,041	13,700			1.00246 $E-08$
18	179,021	15,000			9.09750 $E-09$
19	166,200	15,175	19,262.5	138,856,871	8.98563 $E-09$
20	180,732	14,995			9.10073 $E-09$
21	178,187	15,050			9.06525 $E-09$
22	185,304	15,200	16,650.0	132,388,992	8.96987 $E-09$
23	155,931	15,150			9.00144 $E-09$
24	172,579	16,800			8.06478 $E-09$
25	188,851	16,500	19,262.5	138,856,871	8.22030 $E-09$
26	192,424	17,830			7.57287 $E-09$
27	203,112	19,500			6.89136 $E-09$
28	192,482	19,200	19,262.5	138,856,871	7.00460 $E-09$
29	218,715	19,000			7.08218 $E-09$
30	214,317	19,350			6.94752 $E-09$

Descrição

A regressão de mínimos quadrados de s_y^2 contra \bar{x} levou ao seguinte ajuste:

$$\widehat{s}_y^2 = -9226,002 + 7781,626\bar{x}.$$

Então, o i -ésimo peso será dado por

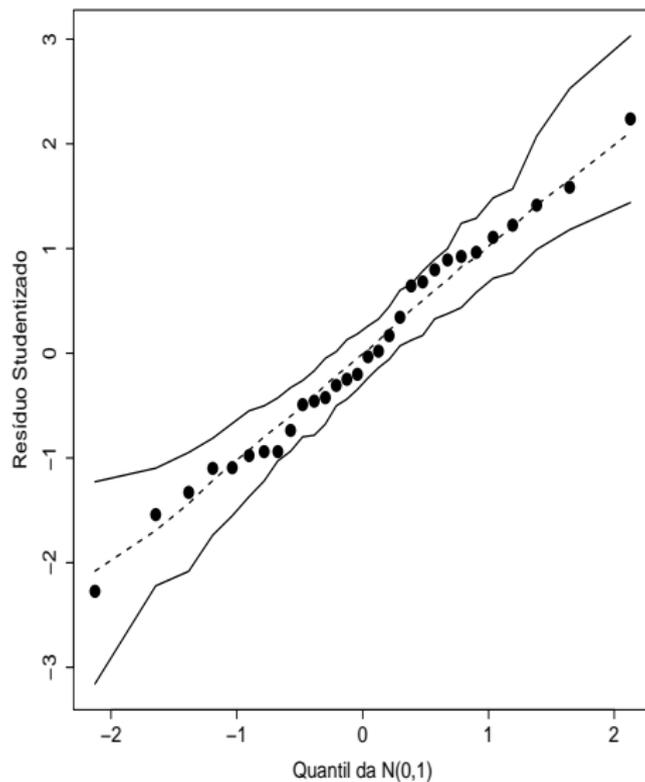
$$\omega_i = -9226,002 + 7781,626x_i,$$

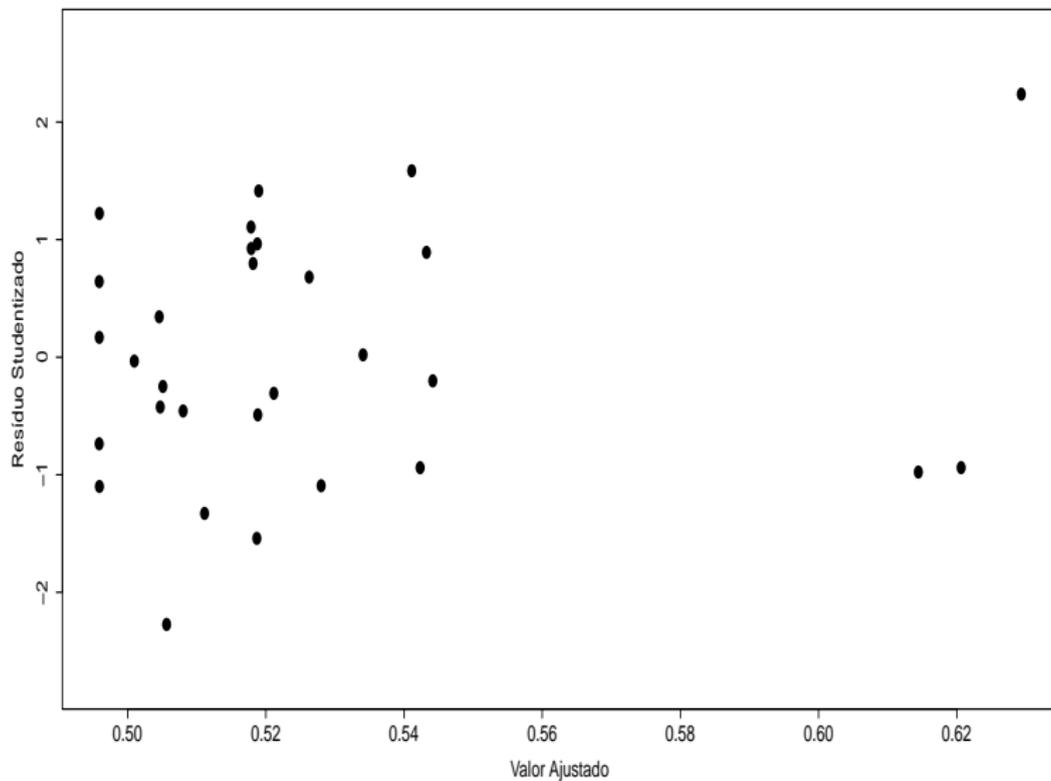
em que x_i denota o gasto com publicidade do i -ésimo restaurante, $i = 1, \dots, 30$.

Descrição das Estimativas

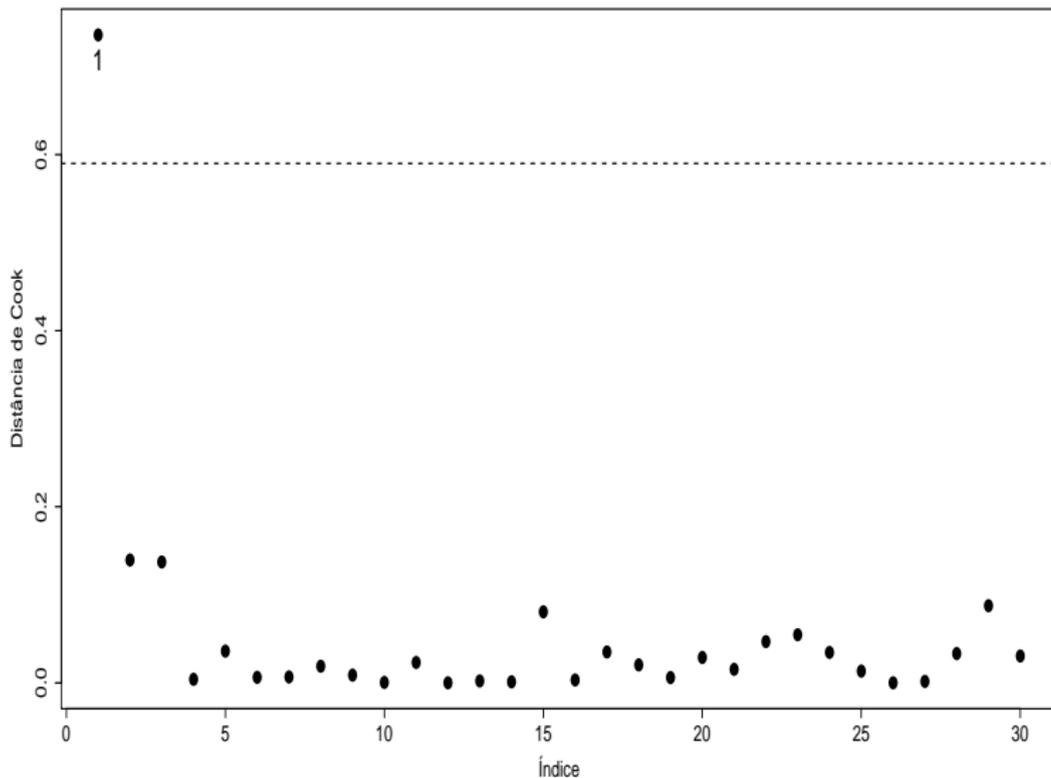
Efeito	Estimativa	E.Padrão	valor-t	valor-P
Constante	51,025	2,440	30,92	0,00
Publicidade	7,918	0,251	31,52	0,00
s	0,031			
R ²	0,997			
\bar{R}^2	0,997			

Os coeficientes são marginalmente significativos.





Distância de Cook



Comentários

O gráfico normal de probabilidades continua indicando um ajuste adequado e há indícios pelo gráfico do resíduo Studentizado contra o valor ajustado de que **a variabilidade foi controlada**. A observação #1 que aparece como supostamente influente não traz mudanças inferenciais e apenas uma variação acima do esperado para a estimativa do intercepto.

- 1 Faturamento de Restaurantes
- 2 Análise de Dados Preliminar
- 3 Regressão Linear Simples
- 4 Regressão Ponderada
- 5 Ajuste Modelo Duplo**
- 6 Conclusões
- 7 Referências

Descrição

Alternativamente, pode-se aplicar modelagem dupla

$$y_i = \beta_1 + \beta_2 \times \text{publicidade}_i + \epsilon_i,$$

para $i = 1, \dots, 30$, em que y_i denota o faturamento anual do i -ésimo restaurante e $\epsilon_i \stackrel{\text{ind}}{\sim} N(0, \sigma_i^2)$ com

$$\log(\sigma_i^2) = \gamma_1 + \gamma_2 \times \text{publicidade}_i,$$

sendo β_1 , β_2 , γ_1 e γ_2 os parâmetros a serem estimados.

Estimativas para a Média

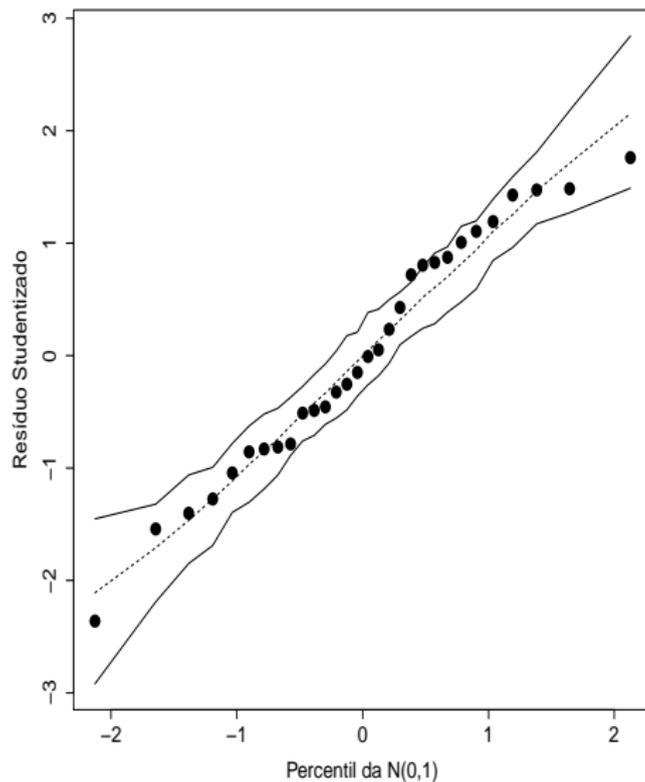
Efeito	Estimativa	E. Padrão	Valor-z	Valor-P
Constante	51,020	2,709	18,83	0,000
Publicidade	7,898	0,273	28,93	0,000

Os coeficientes são marginalmente significativos.

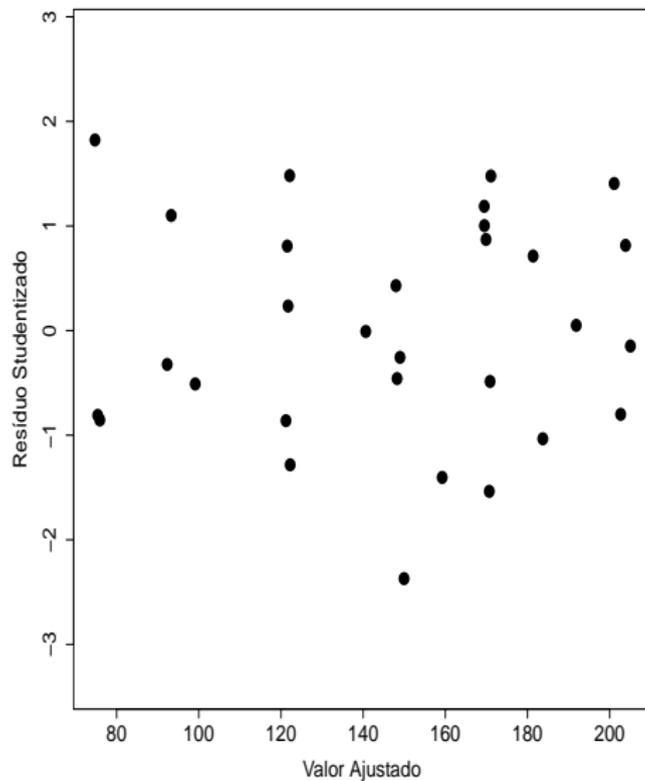
Estimativas para a Variância

Efeito	Estimativa	E. Padrão	Valor-z	Valor-P
Constante	2,516	0,674	3,73	0,000
Publicidade	0,135	0,051	2,65	0,008

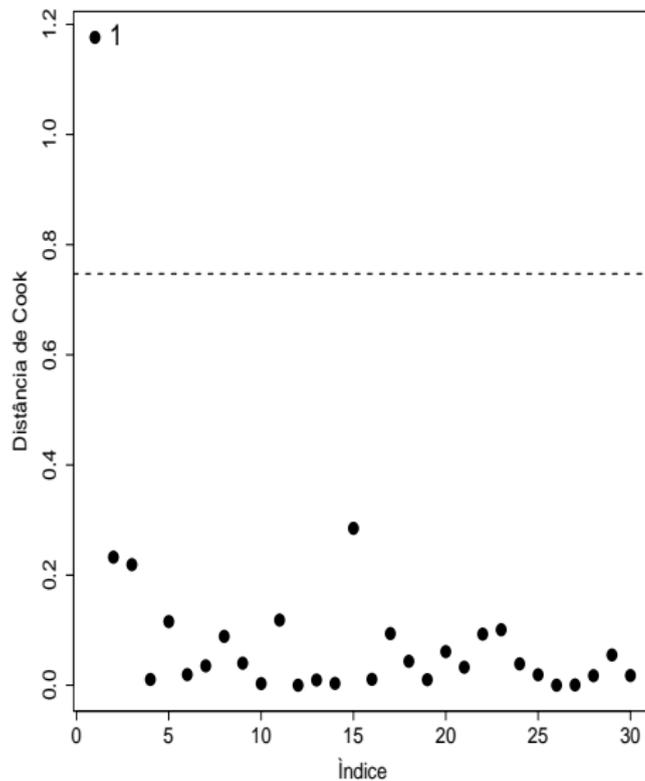
Os coeficientes são marginalmente significativos.

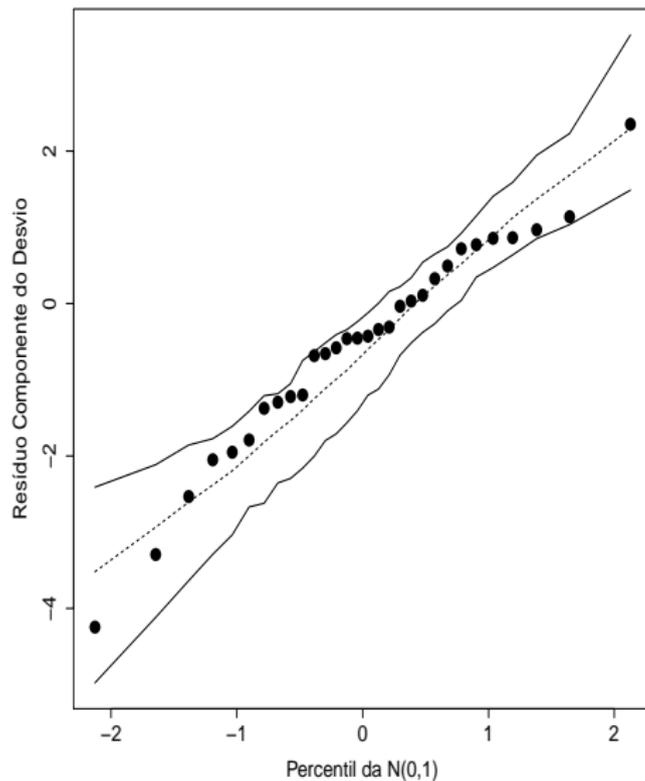


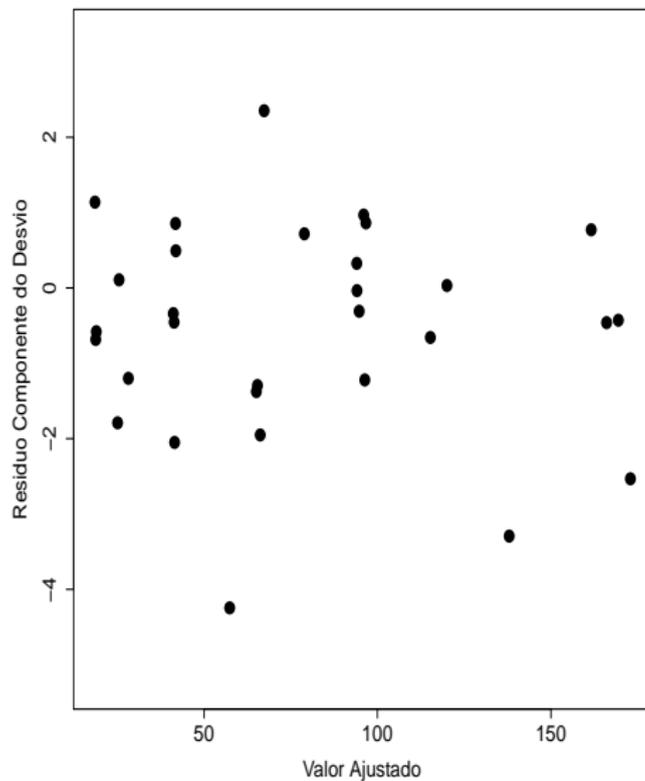
Análise de Resíduos Localização



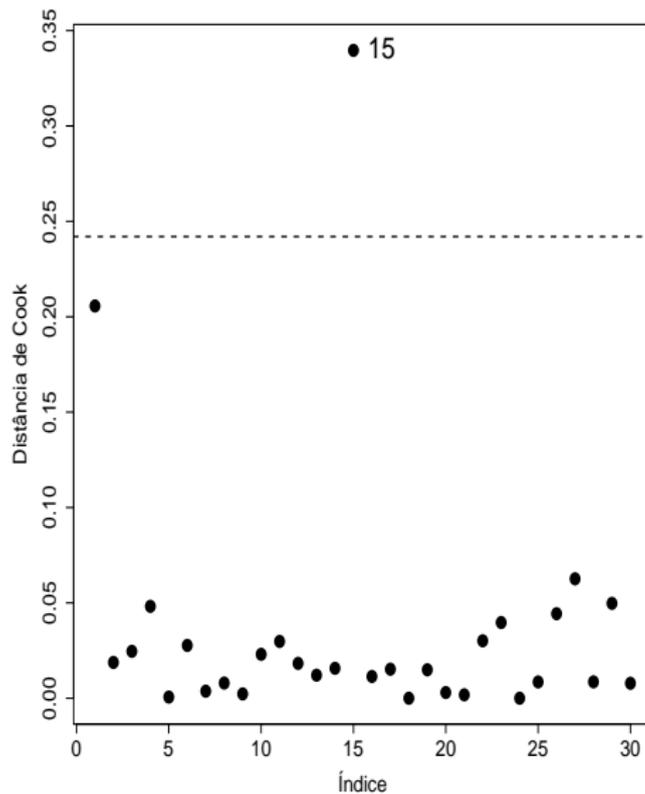
Distância de Cook Localização







Distância de Cook Dispersão



Comentários

- As observações #1 e #15 aparecem como possivelmente influentes nas estimativas dos modelos para a média e para a dispersão, respectivamente.
- A eliminação individual de cada observação não muda a inferência, todos os coeficientes continuam significativos.

Ajuste para a Média

O faturamento predito (dado o gasto com publicidade) fica dado por

$$\hat{\mu}(\text{publicidade}) = 51,020 + 7,898 \times \text{publicidade},$$

em que $\hat{\mu}(\text{publicidade})$ denota o valor predito de faturamento para o gasto com publicidade.

Ajuste para a Variância

A variância predita do faturamento (dado o gasto com publicidade) fica dada por

$$\hat{\sigma}^2(\text{publicidade}) = \exp(2,516 + 0,135 \times \text{publicidade}),$$

em que $\hat{\sigma}^2(\text{publicidade})$ denota o valor predito para a variância do faturamento para o gasto com publicidade.

Estimativas para a Média

Para cada aumento de 1000 USD no gasto anual com publicidade espera-se aumento de 7898 USD no faturamento anual do restaurante.

Estimativas para a Variância

Para cada aumento de 1000 USD no gasto anual com publicidade espera-se aumento de **14,5%** $((\exp(0,135)-1) \times 100\%)$ na variância do faturamento anual.

Estimativas para a Média

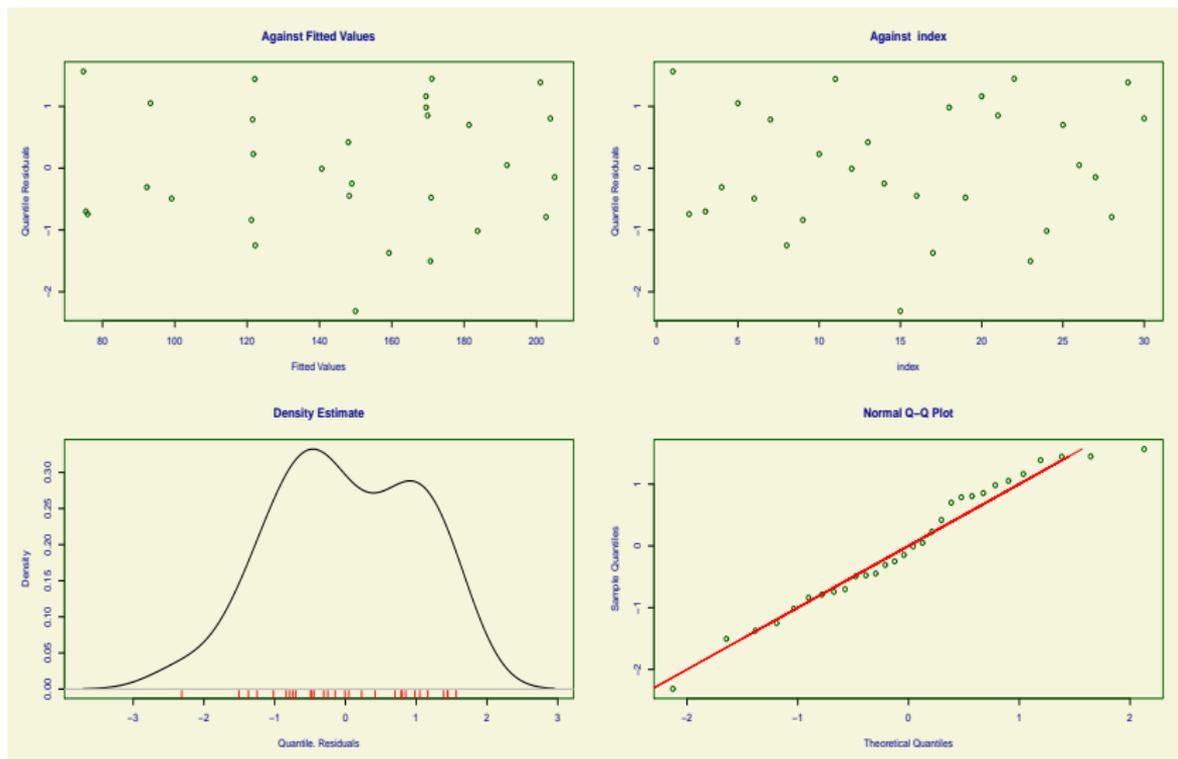
Efeito	Estimativa	E. Padrão	Valor-z	Valor-P
Constante	51,017	2,653	19,23	0,00
Publicidade	7,898	0,268	29,45	0,00

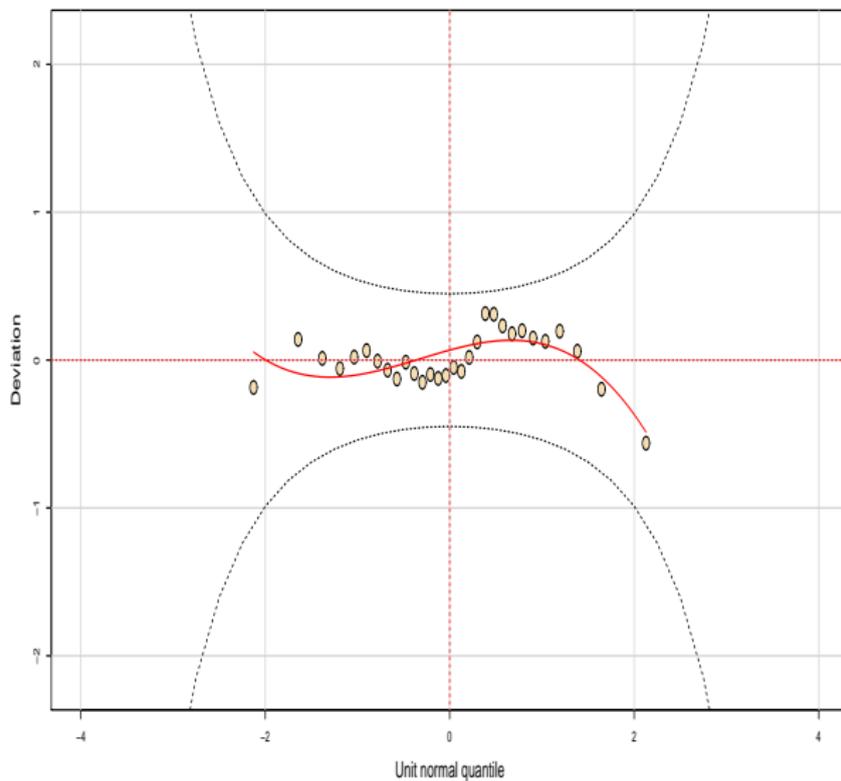
Os coeficientes são marginalmente significativos.

Estimativas para o Desvio Padrão

Efeito	Estimativa	E. Padrão	Valor-z	Valor-P
Constante	1,259	0,362	3,47	0,00
Publicidade	0,068	0,028	2,42	0,02

Os coeficientes são marginalmente significativos.





- 1 Faturamento de Restaurantes
- 2 Análise de Dados Preliminar
- 3 Regressão Linear Simples
- 4 Regressão Ponderada
- 5 Ajuste Modelo Duplo
- 6 Conclusões**
- 7 Referências

Considerações Finais

- Este é um exemplo em que há indícios de heterocedasticidade através do gráfico do resíduo padronizado contra o valor ajustado do modelo homocedástico.
- Os ajustes da regressão ponderada e da regressão dupla levam a ajustes adequados. Contudo, a modelagem dupla permite uma interpretação da variação da variância dado o gasto com publicidade.

- 1 Faturamento de Restaurantes
- 2 Análise de Dados Preliminar
- 3 Regressão Linear Simples
- 4 Regressão Ponderada
- 5 Ajuste Modelo Duplo
- 6 Conclusões
- 7 Referências**

Referências

- Dunn, P. K. e Smith, G. K. (2020). R Package dglm on Double Generalized Linear Models.
- Montgomery, D. C.; Peck, E. A. e Vining, G. G. (2021). *Introduction to Linear Regression Analysis, Fifty Edition*. Hoboken: Wiley.
- Stasinopoulos, M. D., Rigby, R. A., Gillian, Z. A., Voudouris, V. e de Bastiani, F. (2017). *Flexible Regression and Smoothing Using GAMLSS in R*. Chapman and Hall/CRC.