

# Banco de dados: Boston House Prices

Fernado Bispo, Jeff Caponero

# Sumário

|   |    |
|---|----|
| Sobre o banco de dados . . . . .                    | 2  |
| Contexto . . . . .                                  | 2  |
| Objetivo . . . . .                                  | 2  |
| Informações do conteúdo do banco de dados . . . . . | 2  |
| Variável de saída: . . . . .                        | 3  |
| Fonte . . . . .                                     | 3  |
| Análise Descritiva . . . . .                        | 3  |
| Referências . . . . .                               | 10 |

## Sobre o banco de dados

### Contexto

Os dados de preços de 506 casas em Boston publicados em HHarrison, D. and Rubinfeld, D.L. 'Hedonic prices and the demand for clean air', J. Environ. Economics & Management, vol.5, 81-102, 1978.

Os dados podem ser acessados em: <https://www.kaggle.com/datasets/fedesoriano/the-boston-houseprice-data>.

### Objetivo

O objetivo deste trabalho será determinar, a partir de técnicas de regressão linear, o preço de casas em Boston com base nos dados fornecidos pelo banco de dados analisado.

### Informações do conteúdo do banco de dados

- 1) CRIM: índice de criminalidade per capita por bairro.
- 2) ZN: proporção de terreno residencial zoneada para lotes acima de 25.000 sq.ft.
- 3) INDUS: proporção de hectares de negócios não varejistas por bairro.
- 4) CHAS: Margem do rio Charles (1 se o trecho margeia o rio; 0 caso contrário).
- 5) NOX: concentração de óxidos nítricos (partes por 10 milhões) [partes/10M].
- 6) RM: número médio de cômodos por habitação.
- 7) AGE: proporção de unidades próprias construídas antes de 1940.
- 8) DIS: distâncias ponderadas para cinco centros de emprego de Boston.
- 9) RAD: índice de acessibilidade às rodovias radiais.
- 10) TAX: valor total do imposto predial por \$10.000 [\$ / 10k].
- 11) PTRATIO: proporção aluno-professor por bairro.

12) B: O resultado da equação  $B = 1000(Bk - 0,63)^2$  onde  $Bk$  é a proporção de negros por bairro.

13) LSTAT: % da população de “classe baixa”.

### Variável de saída:

1) MEDV: Valor médio de residências ocupadas pelo proprietário em US\$1.000 [k\$].

### Fonte

StatLib - Carnegie Mellon University

## Análise Descritiva

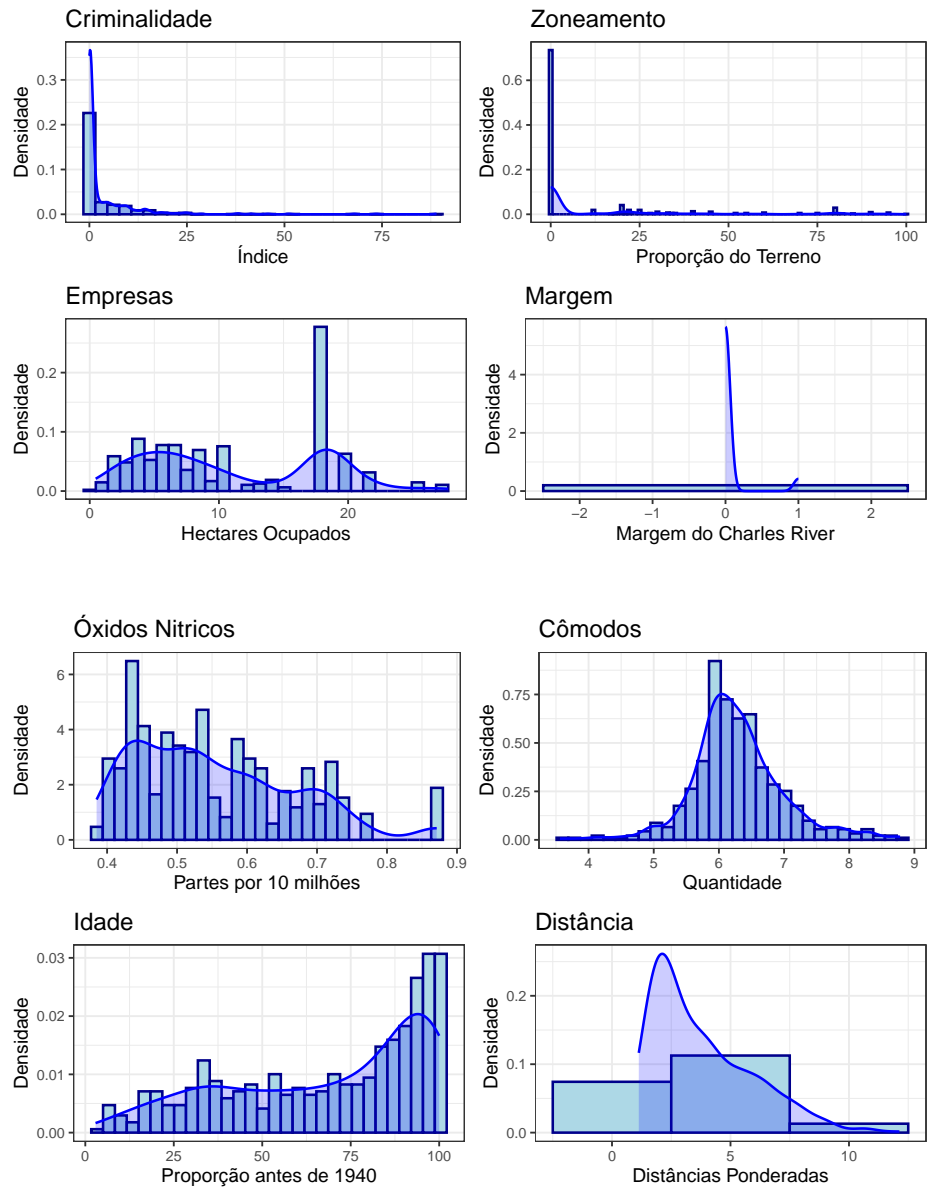
Table 1: Medidas Resumo dos dados

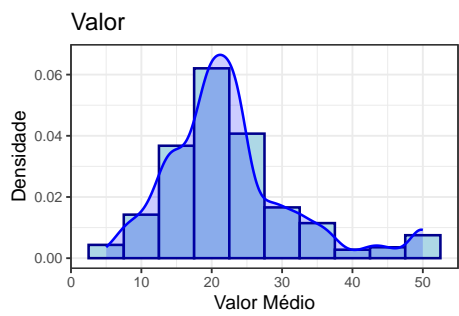
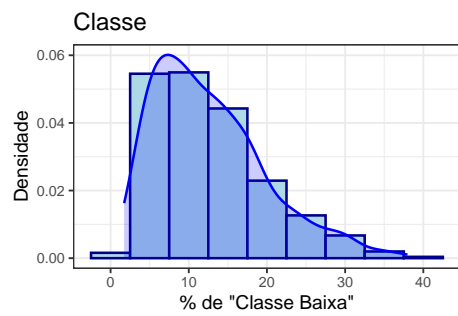
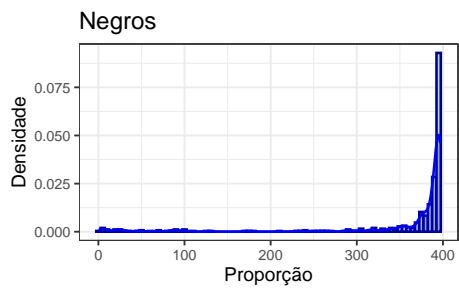
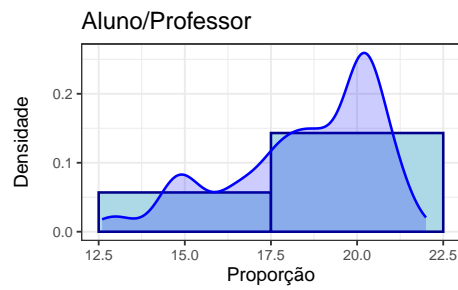
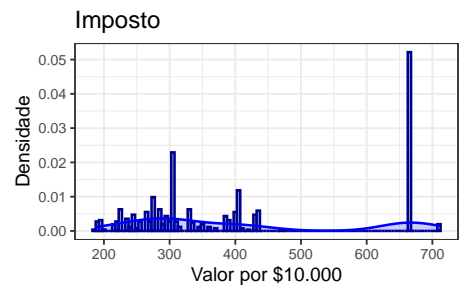
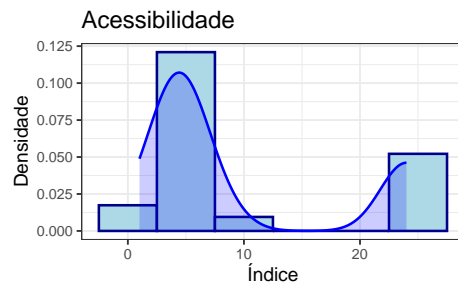
|         | Min    | Q1     | Med    | Média  | Q3     | Max    | D.Padrão | CV   |
|---------|--------|--------|--------|--------|--------|--------|----------|------|
| AGE     | 2,90   | 45,00  | 77,50  | 68,57  | 94,10  | 100,00 | 28,15    | 0,41 |
| B       | 0,32   | 375,33 | 391,44 | 356,67 | 396,23 | 396,90 | 91,29    | 0,26 |
| CHAS    | 0,00   | 0,00   | 0,00   | 0,07   | 0,00   | 1,00   | 0,25     | 3,67 |
| CRIM    | 0,01   | 0,08   | 0,26   | 3,61   | 3,68   | 88,98  | 8,60     | 2,38 |
| DIS     | 1,13   | 2,10   | 3,21   | 3,80   | 5,21   | 12,13  | 2,11     | 0,55 |
| INDUS   | 0,46   | 5,19   | 9,69   | 11,14  | 18,10  | 27,74  | 6,86     | 0,62 |
| LSTAT   | 1,73   | 6,93   | 11,36  | 12,65  | 16,96  | 37,97  | 7,14     | 0,56 |
| MEDV    | 5,00   | 17,00  | 21,20  | 22,53  | 25,00  | 50,00  | 9,20     | 0,41 |
| NOX     | 0,38   | 0,45   | 0,54   | 0,55   | 0,62   | 0,87   | 0,12     | 0,21 |
| PTRATIO | 12,60  | 17,40  | 19,05  | 18,46  | 20,20  | 22,00  | 2,16     | 0,12 |
| RAD     | 1,00   | 4,00   | 5,00   | 9,55   | 24,00  | 24,00  | 8,71     | 0,91 |
| RM      | 3,56   | 5,88   | 6,21   | 6,28   | 6,62   | 8,78   | 0,70     | 0,11 |
| TAX     | 187,00 | 279,00 | 330,00 | 408,24 | 666,00 | 711,00 | 168,54   | 0,41 |
| ZN      | 0,00   | 0,00   | 0,00   | 11,36  | 12,50  | 100,00 | 23,32    | 2,05 |

*Note:*

Fonte: StatLib - Carnegie Mellon University

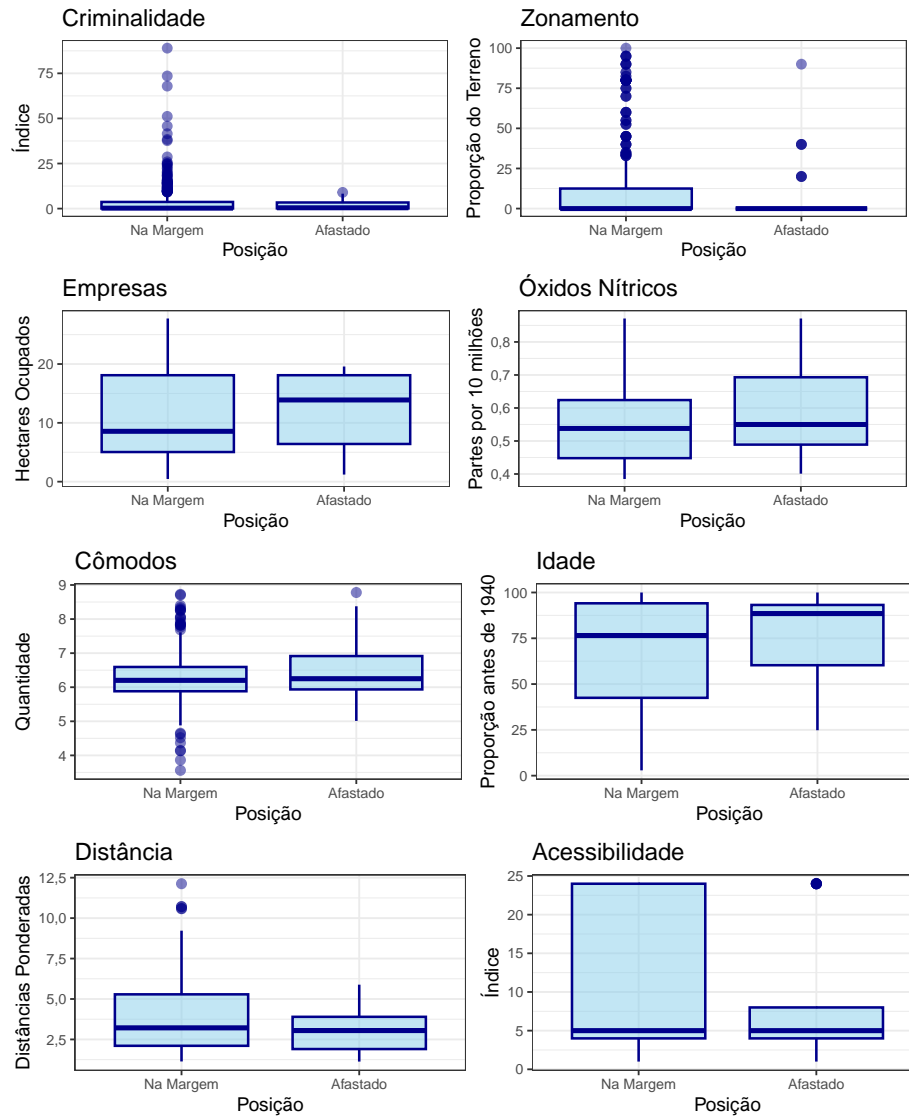
Figura 2: Histogramas das variáveis em análise.

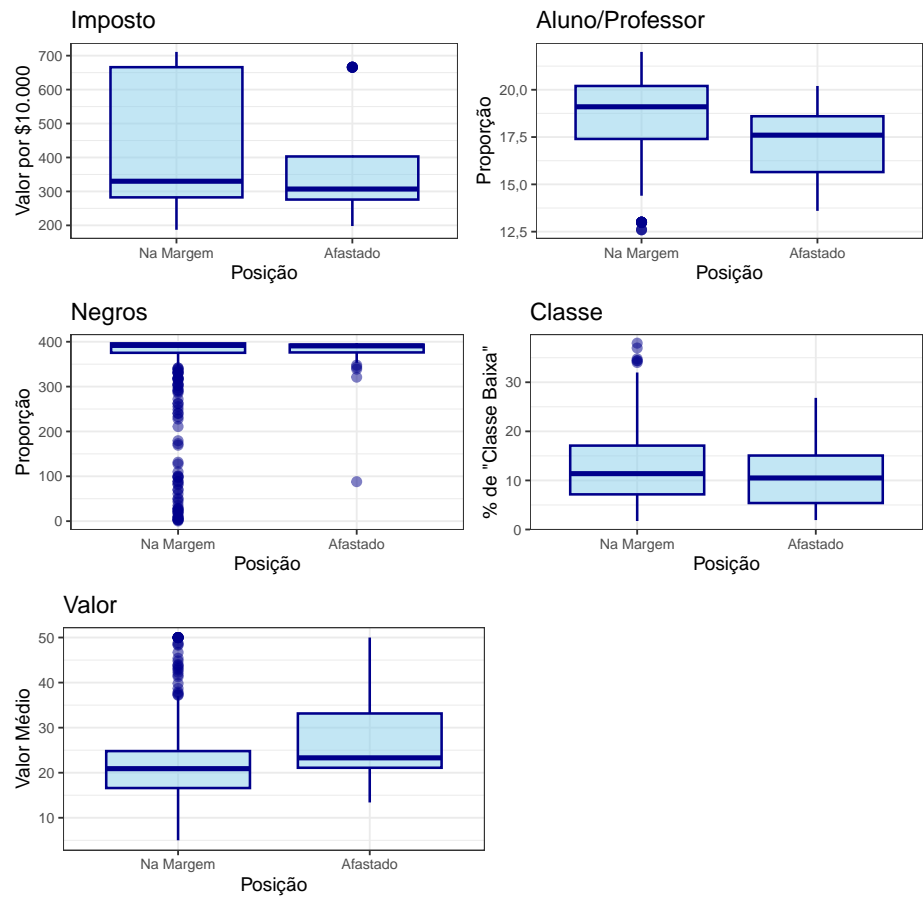




Fonte: StatLib – Carnegie Mellon University

Figura 3: BoxPlots entre a posição em relação ao Charles River e demais variáveis em :

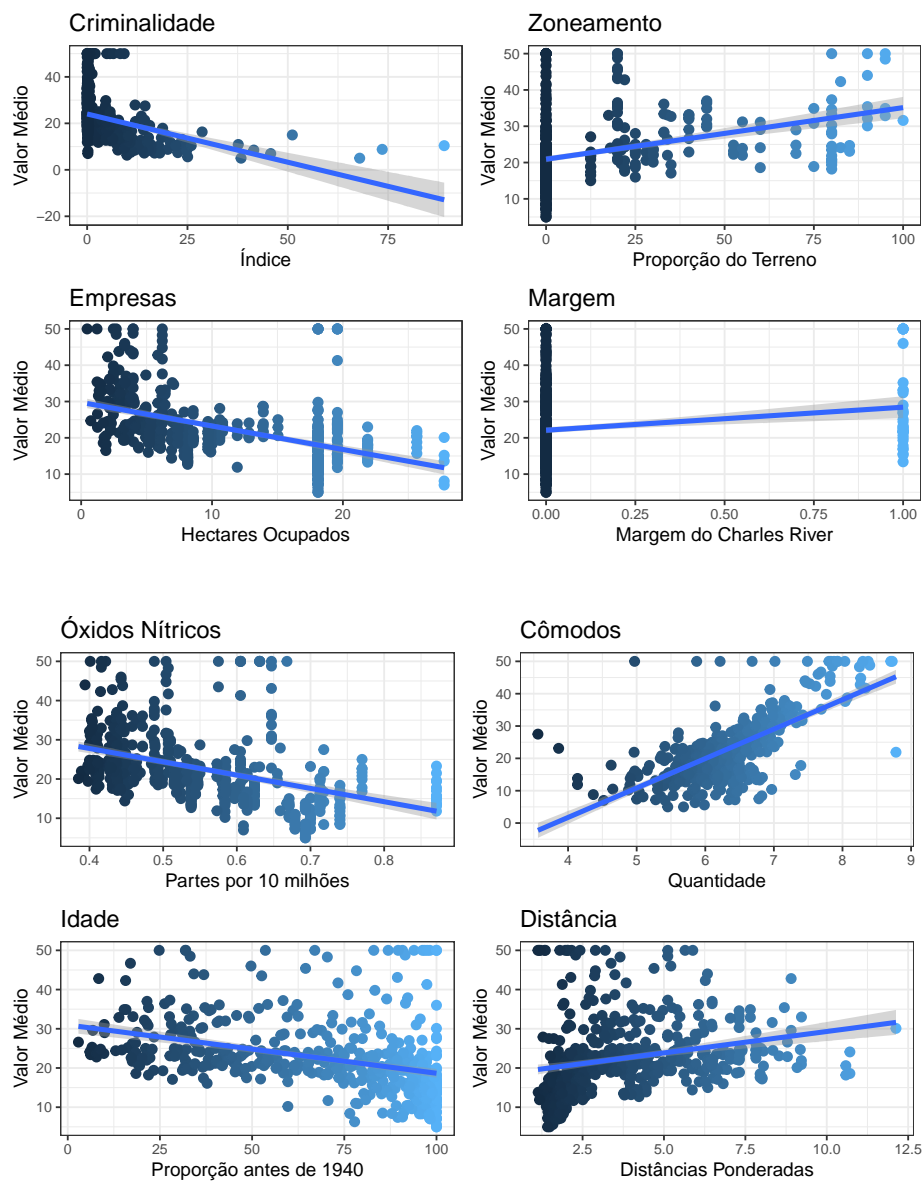


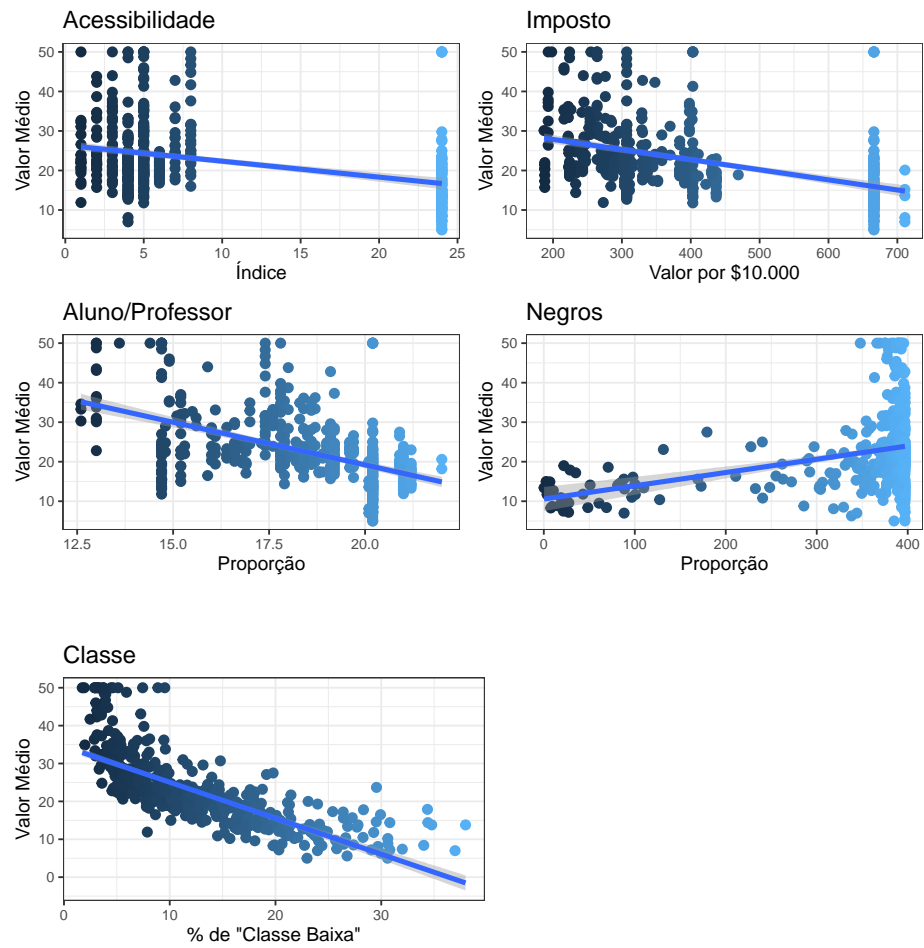


Fonte: StatLib – Carnegie Mellon University



Figura 4: Relação entre a Função Diabética e demais medições





Fonte: StatLib – Carnegie Mellon University

Table 2: Valores dos modelos de regressão linear simples.

|             | CRIM  | ZN    | INDUS | CHAS  | NOX    | RM     |
|-------------|-------|-------|-------|-------|--------|--------|
| beta0       | 24,03 | 20,92 | 29,75 | 22,09 | 41,35  | -34,67 |
| sigma0      | 0,41  | 0,42  | 0,68  | 0,42  | 1,81   | 2,65   |
| beta1       | -0,42 | 0,14  | -0,65 | 6,35  | -33,92 | 9,10   |
| sigma1      | 0,04  | 0,02  | 0,05  | 1,59  | 3,20   | 0,42   |
| p-valor     | 0,00  | 0,00  | 0,00  | 0,00  | 0,00   | 0,00   |
| Coef. Corr. | 0,15  | 0,13  | 0,23  | 0,03  | 0,18   | 0,48   |

|             | AGE   | DIS   | RAD   | TAX   | PTRATIO | B     | LSTAT |
|-------------|-------|-------|-------|-------|---------|-------|-------|
| beta0       | 30,98 | 18,39 | 26,38 | 32,97 | 62,34   | 10,55 | 34,55 |
| sigma0      | 1,00  | 0,82  | 0,56  | 0,95  | 3,03    | 1,56  | 0,56  |
| beta1       | -0,12 | 1,09  | -0,40 | -0,03 | -2,16   | 0,03  | -0,95 |
| sigma1      | 0,01  | 0,19  | 0,04  | 0,00  | 0,16    | 0,00  | 0,04  |
| p-valor     | 0,00  | 0,00  | 0,00  | 0,00  | 0,00    | 0,00  | 0,00  |
| Coef. Corr. | 0,14  | 0,06  | 0,15  | 0,22  | 0,26    | 0,11  | 0,54  |

## Referências

Harrison, David & Rubinfeld, Daniel. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*. 5. 81-102. 10.1016/0095-0696(78)90006-2.

Belsley, David A. & Kuh, Edwin. & Welsch, Roy E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. New York: Wiley.