

# Precificação de apartamentos para o bairro Água Verde em Curitiba

Chuck Norris\*

Arnold Schwarzenegger†

6 de dezembro de 2013

O preço de imóveis depende principalmente do seu tamanho e localização. A infraestrutura dos bairros influenciam o preço do imóvel pela disponibilidade de recursos, como mercados e escolas, vias de acessos e distâncias com relação aos centros comerciais da cidade, além de segurança e arborização. Dentro de um mesmo bairro o valor dos imóveis dependem da sua área construída, da face voltada para o sol, do número de quartos e das vagas na garagem. Diante de tantas características que regulam o valor do imóvel é útil estabelecer um modelo para precificação a partir de suas características mais relevantes. O objetivo com essa análise é construir um modelo para precificação de apartamentos no bairro Água Verde em Curitiba.

## 1 Descrição e análise exploratória

Uma amostra aleatória de 40 apartamentos tipo padrão foi feita dentro do site [Imóveis Curitiba](#). Coletou-se apenas informações para imóveis do bairro Água Verde. Foram registradas as variáveis do imóvel: preço (R\$), área (m<sup>2</sup>), número de quartos, de suítes, de banheiros e vagas na garagem. Os dados estão disponíveis na tabela 3 na sessão anexos.

A análise exploratória dos dados aponta que o valor do imóvel aumenta com a área do mesmo (figura 1). Na escala original das variáveis preço e área, verifica-se uma relação linear positiva com aumento da variabilidade ao longo da linha de tendência ao passo que na escala logarítmica essas duas variáveis apresentam ainda uma relação linear porém com variabilidade aproximadamente constante.

Da amostra coletada mais de 50% dos apartamentos têm até dois banheiros (tabela 1). Mais da metade dos apartamentos têm três quartos e quase 90% deles têm uma suíte e até duas vagas na garagem. O preço do imóvel aumenta de acordo com o número de banheiros, quartos, suítes e vagas na garagem conforme pode ser visto na figura 2. No entanto, não pode deixar de ser mencionado que o aumento do número de quartos, por exemplo, também representa aumento da área do imóvel. Sendo assim, as relações observadas nos gráficos não podem ser interpretadas isoladamente pois essas variáveis que regulam o preço do imóvel são dependentes umas das outras.

Tabela 1: Frequência do número banheiros, quartos, suítes e vagas na garagem por imóvel. Valores entre parenteses representam a frequência relativa.

unidades	banheiros	quartos	suítes	vagas
1	10 (25.0%)	1 (2.5%)	35 (87.5%)	16 (40.0%)
2	17 (42.5%)	10 (25.0%)	3 (7.5%)	19 (47.5%)
3	8 (20.0%)	23 (57.5%)	2 (5.0%)	4 (10.0%)
4	3 (7.5%)	6 (15.0%)		1 (2.5%)
5	1 (2.5%)			
6	1 (2.5%)			

## 2 Ajuste do modelo

Com a intenção de precificar imóveis será ajustado um modelo de regressão linear simples do preço do imóvel como função da área. Os valores serão usados na escala logarítmica de base 10 uma vez que o

\*Acadêmico do Curso de Estatística, grr: 12345678.

†Acadêmico do Curso de Estatística, grr: 87654321.

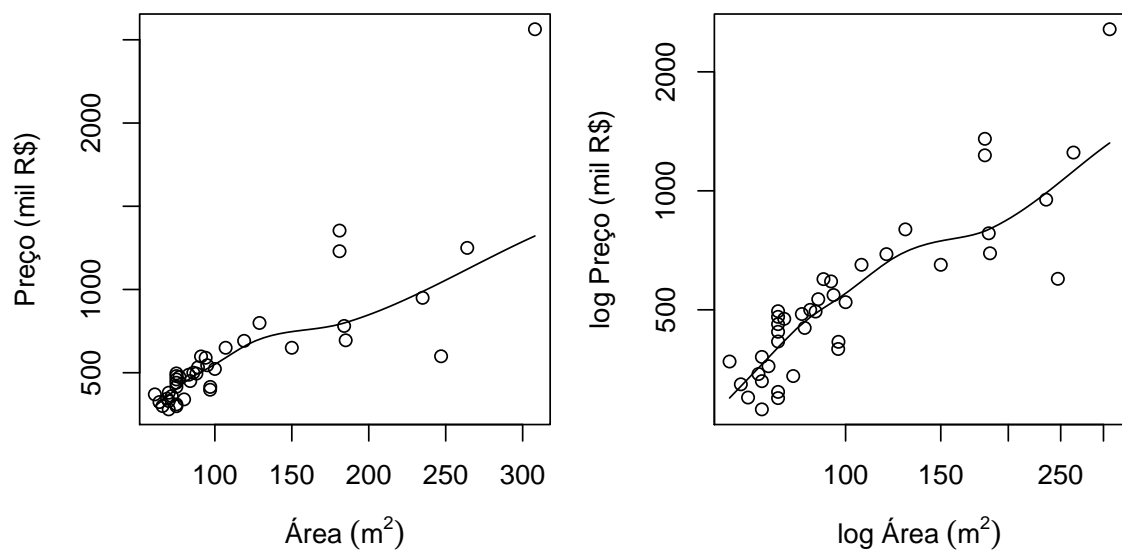


Figura 1: Preço do imóvel (R\$) em função da sua área (m<sup>2</sup>) na escala natural (esq.) e na escala logarítmica dos valores (dir.). Linha suave indica a relação de tendência entre as variáveis.

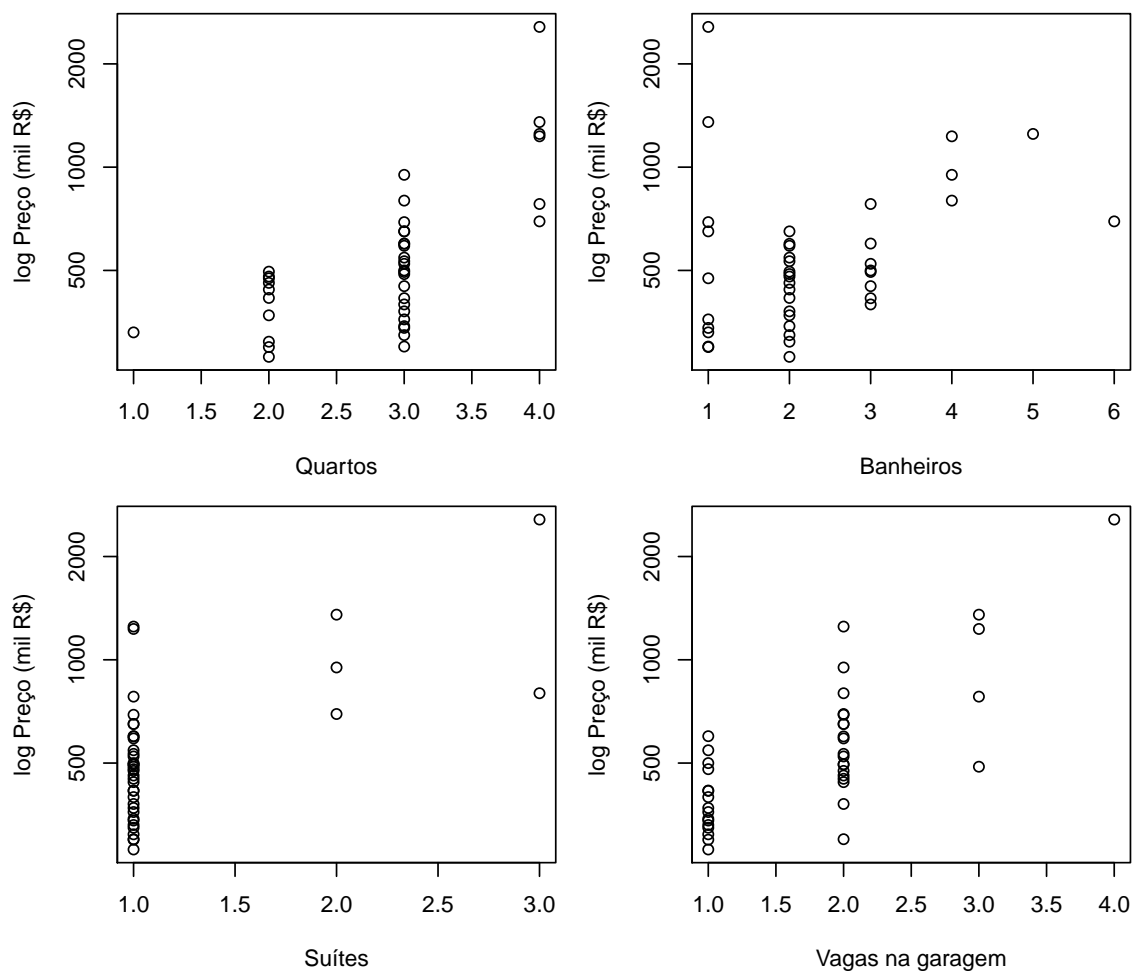


Figura 2: Preço do imóvel (R\$) em função do número de quartos, banheiros, suítes e vagas na garagem.

gráfico de análise exploratória indicou uma relação linear de variabilidade constante para a relação entre essas duas variáveis.

O modelo assumido para os dados é representado por

$$Y|x \sim \text{Normal}(\eta(x, \boldsymbol{\beta}), \sigma^2), \quad (1)$$

$$\eta(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x, \quad (2)$$

em que  $Y|x$  representa a distribuição condicional do preço dos imóveis aos valores de sua área construída. A média de preço dos imóveis é uma função linear ( $\eta$ ) de sua área construída e a variação de preço ao redor do valor médio tem distribuição normal com variância constante ( $\sigma$ ), ou seja, que independe do próprio valor, da área ou qualquer outra característica.

A estimação dos parâmetros é feita por máxima verossimilhança que consiste em encontrar o máximo da função de verossimilhança

$$L(\beta_0, \beta_1, \sigma^2; y, x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y_i - (\beta_0 + \beta_1 x_i))^2}{2\sigma^2} \right\}, \quad (3)$$

em relação aos parâmetros do modelo, no caso,  $\beta_0$ ,  $\beta_1$  e  $\sigma^2$ . Como já mencionado, será usando o log do preço e da área pois nessa escala tem-se uma maior compatibilidade com dois dos pressupostos do modelo: relação linear e variância constante.

```
imo <- transform(imo, lp=log10(preco), la=log10(area))
m0 <- lm(lp~la, data=imo)
summary(m0)

##
## Call:
## lm(formula = lp ~ la, data = imo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3091 -0.0723  0.0124  0.0625  0.2324
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.8420     0.1687    4.99 1.4e-05 ***
## la            0.9381     0.0837   11.20 1.3e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0999 on 38 degrees of freedom
## Multiple R-squared:  0.768, Adjusted R-squared:  0.761
## F-statistic: 125 on 1 and 38 DF, p-value: 1.33e-13
```

O ajuste do modelo aos dados forneceu a seguinte equação estimada para o valor esperado do imóvel na escala  $\log_{10}$

$$E(\log_{10}(\text{preço})|\text{área}) = 0.842 + 0.938 \log_{10}(\text{área}), \quad (4)$$

com coeficiente de determinação de  $R^2 = 0.768$ . Pelas estimativas dos parâmetros tem-se que o valor do imóvel aumenta significativamente com a área ( $p$ -valor  $< 0.01$ ). No entanto, tais conclusões só são válidas mediante o atendimento dos pressupostos do modelo considerado. Os pressupostos são: 1) que o modelo considerado representa a relação entre as variáveis, 2) que os desvios têm variabilidade constante ao redor dos valores dados pelo modelo ajustado, 3) que os desvios têm distribuição normal e que 4) as observações são independentes. O requisito 4) é justificado pelo processo aleatório de aquisição dos dados. Para os demais, análise gráfica dos resíduos fornecem uma boa idéia sobre o atendimento dos pressupostos.

Pela figura 3 observa-se não haver fuga do pressuposto 1) pois os resíduos se distribuem ao redor da linha horizontal do zero. Para o pressuposto 2) também não se verifica afastamento de pressuposto uma vez que os resíduos ocupam uma faixa de mesma altura ao longo do eixo dos valores preditos, indicando haver nenhuma ou muito fraca relação da variabilidade com a o valor ajustado pelo modelo. O pressuposto de normalidade, 3), também é atendido pois os pontos formam uma linha diagonal. Portanto, o modelo ajustado, uma vez que os pressupostos foram atendidos, pode ser considerado para inferência e previsão de preços.

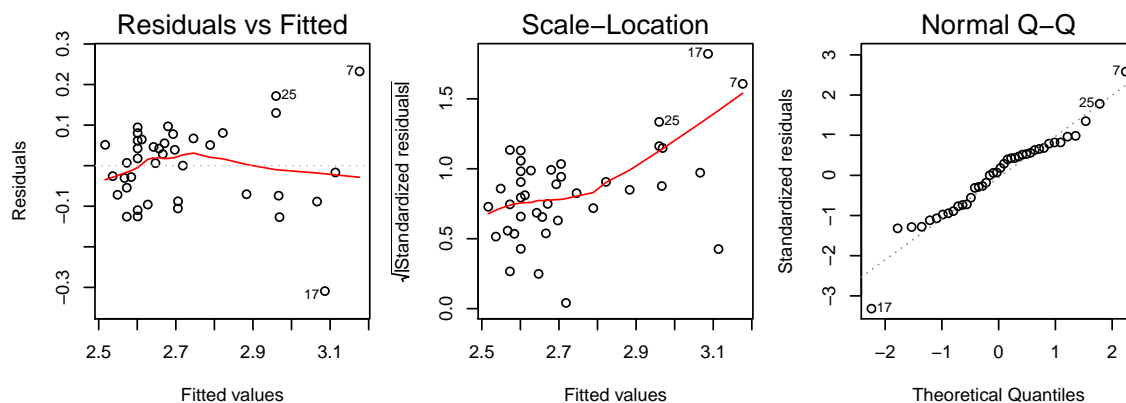


Figura 3: Gráficos de análise dos resíduos. Resíduos em função dos valores ajustados (esquerda). Raiz quadrada dos resíduos padronizados em função dos valores ajustados (centro). Quantis observados em função dos quantis teóricos da distribuição normal (direita).

A figura 4 tem-se os valores observados e os ajustados dados pelo modelo. As bandas ao redor dos valores preditos delimitam o intervalo de confiança (95%) para cada valor predito. Os valores sobre a linha são portanto os preços esperados para os apartamentos considerando sua área. Na tabela 2 estão os valores preditos para apartamentos de cinco tamanhos.

Tabela 2: Preço predito do imóvel em função da sua área. LI e LS são o limite inferior e superior do intervalo de confiança (95%) para o valor predito.

Área (m <sup>2</sup> )	$\log_{10}$ Área	Preço (mil R\$)	LI	LS
50	1.70	272.75	236.99	313.90
75	1.88	398.98	364.81	436.35
100	2.00	522.58	485.48	562.52
150	2.18	764.43	692.22	844.18
200	2.30	1001.25	873.23	1148.04

### 3 Considerações finais

Diante das várias características que regulam o preço do imóvel estabeleceu-se um modelo baseado apenas na área do mesmo. O modelo de regressão linear simples, considerando o logaritmo do preço do imóvel como função do logaritmo da área (ambos em base 10), ajustado atendeu os pressupostos e apresentou um coeficiente de determinação de 0.768. Com o modelo foi possível estimar o preço de apartamentos a partir da área localizados no bairro Água Verde em Curitiba.

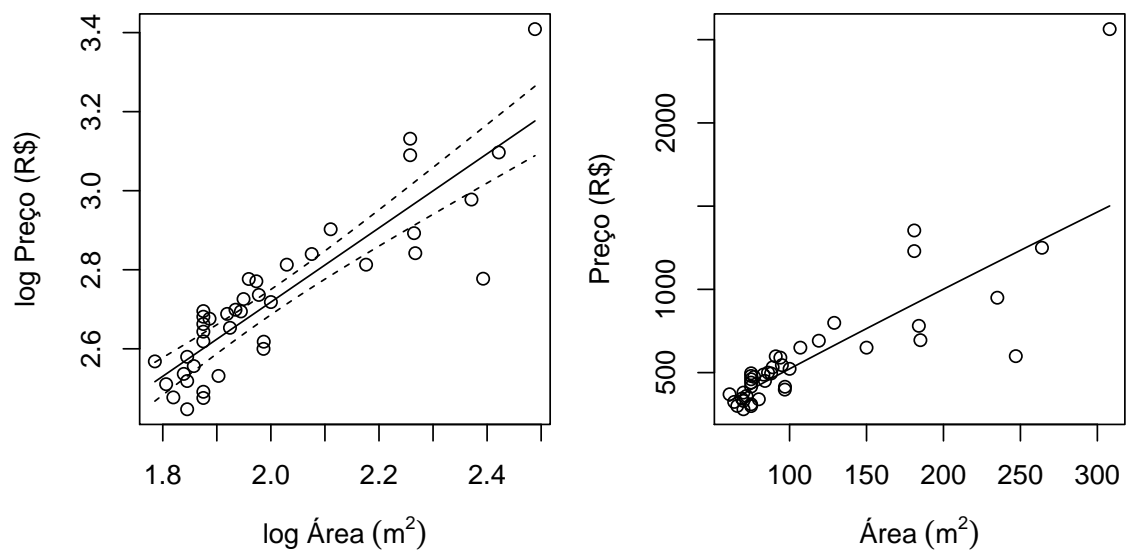


Figura 4: Valores observados (pontos) e ajustados (linha) para preço em função da área do imóvel na escala logarítmica e na escala natural.

## 4 Anexos

Tabela 3: Variáveis para uma amostra aleatória de 40 apartamento no bairro Água Verde.

Preço (R\$)	Área (m <sup>2</sup> )	Quartos	Banheiros	Vagas	Suítes
280.00	70.0	2	2	1	1
299.00	75.0	2	1	1	1
300.00	66.0	3	1	2	1
310.00	75.0	2	2	1	1
324.00	64.0	3	2	1	1
330.00	70.0	1	1	1	1
340.00	80.0	3	1	1	1
344.04	69.0	3	2	1	1
360.00	72.0	3	1	1	1
370.00	61.0	2	2	1	1
380.00	70.0	3	2	2	1
398.00	96.9	3	3	1	1
415.00	97.0	3	3	1	1
415.80	75.0	2	2	1	1
440.00	75.0	2	2	2	1
450.00	84.0	3	3	2	1
460.00	75.0	2	2	2	1
474.40	77.0	2	1	2	1
480.00	75.0	2	2	1	1
488.00	83.0	3	2	3	1
495.00	88.0	3	3	2	1
496.00	75.0	2	2	2	1
500.00	86.0	3	3	1	1
522.78	100.0	3	3	2	1
532.00	89.0	3	2	2	1
545.00	95.0	3	2	1	1
590.00	94.0	3	2	2	1
598.00	91.0	3	2	2	1
599.00	247.0	3	3	1	1
650.00	107.0	3	2	2	1
650.00	150.0	3	1	2	1
691.60	119.0	3	1	2	1
695.00	185.0	4	6	2	2
781.00	184.0	4	3	3	1
799.00	129.0	3	4	2	3
950.00	235.0	3	4	2	2
1230.00	181.0	4	4	3	1
1250.00	264.0	4	5	2	1
1353.99	181.0	4	1	3	2
2563.51	308.0	4	1	4	3