Avaliação de imóveis a partir de dados de oferta

A Preprint

Luiz Fernando Palin Droubi
SPU/SC
Secretaria de Coordenação e Governança do Patrimônio da União
Florianópolis, SC
lfpdroubi@gmail.com

12 de junho de 2021

Abstract

Um dos grandes problemas práticos da Engenharia de Avaliações no Brasil é a falta de dados de mercado envolvendo transações de imóveis para elaboração de avaliações com o Método Comparativo de Dados de Mercado. É usual, portanto, a utilização de dados de oferta para a confecção de modelos estatísticos, visando utilizá-la como parâmetro para a aferição do valor de venda. Este trabalho pretende demonstrar diversas formas para o tratamento destes dados de oferta, buscando um conjunto de procedimentos que propriciem a obtenção de resultados consistentes, qualquer quer seja o método aplicado.

***K*eywords** Fator Oferta · Campo de Arbítrio · Intervalos admissíveis

1 Introdução

2 Revisão Bibliográfica

Segundo Matloff [(2009), p. 54), a variância ($\mathbb{V}$) do produto de uma variável aleatória (U) por uma constante (cU) é igual à variância desta mesma variável aleatória U, multiplicada pela constante ao quadrado (c^2), *i.e.*:

$$\mathbb{V}(cU) = c^2\mathbb{V}(U) \tag{1}$$

Isto tem implicações na Engenharia de Avaliações, já que, apesar desta estar interessada na estimação da variável Valor de Venda (V_{venda}), nem sempre esta variável aleatória pode ser diretamente observada, sendo usual a utilização de dados de oferta, mais facilmente encontrados, ou seja, trabalha-se com o observação da variável V_{oferta} e presume-se que:

$$V_{venda} \approx cV_{oferta} \tag{2}$$

Onde, nos casos práticos, $c \leq 1,0$.

Ocorre que na Engenharia de Avaliações existem ao menos duas abordagens práticas:

1. A aplicação de um fator de redução aos valores de oferta (homogeneização) antes do ajuste de um modelo estatístico com a variável.
2. A utilização da variável V_{oferta} para o ajuste de modelos estatísticos e, com o modelo ajustado, a aplicação de um fator de redução (campo de arbítrio) para a transformação do valor previsto pelo modelo estatístico em um valor de venda, que se deseja prever.

Para a obtenção da estimativa de valor central para os valores de venda a partir da segunda abordagem, basta a multiplicação do valor estimado com o modelo de ofertas, haja vista que [Matloff (2009), p. 47):

$$\mathbb{E}[cU] = c\mathbb{E}[U] \tag{3}$$

Em outros termos, a estimativa de valor central para o valor de venda ($\hat{V}_{venda}$) pode ser calculado de acordo com a equação 4:

$$\begin{aligned}
\hat{V}_{venda} &= \mathbb{E}[V_{venda}] \\
&= \mathbb{E}[cV_{oferta}] \\
&= c\mathbb{E}[V_{oferta}] \\
\hat{V}_{venda} &= c\hat{V}_{oferta}
\end{aligned} \tag{4}$$

Ocorre que, se estima-se a variável V_{venda} através da V_{oferta} como na equação 4, pela propriedade expressa na equação 1, tem-se que:

$$\hat{\mathbb{V}}(\hat{V}_{venda}) = \hat{\mathbb{V}}(c\hat{V}_{oferta}) \tag{5}$$

$$= c^2\hat{\mathbb{V}}(\hat{V}_{oferta}) \tag{6}$$

Deve-se lembrar, portanto, que os intervalos de confiança fazem uso da variância estimada ($\hat{\mathbb{V}}$) para o cálculo dos intervalos de confiança e predição dos coeficientes e dos valores previstos pelo modelo. O intervalo de confiança para a previsão de valores médios à partir de um modelo de regressão linear ajustado para uma variável dependente y pode ser calculado através da equação 7 (Faraway 2004, 41–42):

$$IC = \hat{y} \pm t_{(1-\alpha/2;n-k-1)} s.e.(\hat{y}) \tag{7}$$

Onde $s.e.(\hat{y}) = \sqrt{\hat{\mathbb{V}}(\hat{y})}$ (Matloff 2009, 272).

Isto não ocorreria caso o modelo estatístico fosse elaborado com a variável homogeneizada (V_{venda}), pois os intervalos de confiança construídos por tal modelo já estariam utilizando a variância adequada para o seu cômputo:

$$[\hat{V}_{venda} - t_{(1-\alpha/2;n-k-1)} s.e.(\hat{V}_{venda}); \hat{V}_{venda} + t_{(1-\alpha/2;n-k-1)} s.e.(\hat{V}_{venda})] \tag{8}$$

Onde $s.e.(\hat{V}_{venda}) = \sqrt{\hat{\mathbb{V}}(\hat{V}_{venda})}$ é o erro-padrão do estimador obtido com o modelo ajustado com a variável homogeneizada.

Porém, quando se ajusta o modelo com a variável não homogeneizada (V_{oferta}) e se obtem o valor de venda a partir da aplicação da equação 4, fazendo uso do Campo de Arbítrio do avaliador, deve-se ter em conta que o intervalo de confiança para a variável V_{venda} deve ser calculado com a variância da variável $\hat{V}_{venda}$ e não com a variável $\hat{V}_{oferta}$ obtida com os valores ajustados do modelo.

Pois quando ajusta-se o modelo com os dados de oferta, obtem-se um intervalo de confiança ajustado para valores de oferta, e não para valores de venda, que é o que se deseja estimar:

$$IC_{oferta} = [\hat{V}_{oferta} - t_{(1-\alpha/2;n-k-1)} s.e.(\hat{V}_{oferta}); \hat{V}_{oferta} + t_{(1-\alpha/2;n-k-1)} s.e.(\hat{V}_{oferta})] \tag{9}$$

Para a obtenção do intervalo de confiança para a variável V_{venda} deve-se aplicar as transformações apropriadas, *i.e.* $\hat{V}_{venda} = c\hat{V}_{oferta}$ e $s.e.(\hat{V}_{venda}) = c \cdot s.e.(\hat{V}_{oferta})$, já que o erro padrão do estimador $\hat{V}_{venda}$ deve ser assim derivado:

$$\begin{aligned} s.e.(\hat{V}_{venda}) &= \sqrt{\hat{\mathbb{V}}(\hat{V}_{venda})} \\ &= \sqrt{\hat{\mathbb{V}}(c\hat{V}_{oferta})} \\ &= \sqrt{c^2\hat{\mathbb{V}}(\hat{V}_{oferta})} \\ &= c\sqrt{\hat{\mathbb{V}}(\hat{V}_{oferta})} \end{aligned}$$

$$s.e.(\hat{V}_{venda}) = c \cdot s.e.(\hat{V}_{oferta}) \tag{10}$$

De maneira que pode-se escrever o intervalo de confiança para a estimação do Valor de venda em termos do Valor de oferta como segue:

$$IC_{venda} = [c\hat{V}_{oferta} - t_{(1-\alpha/2;n-k-1)}c \cdot s.e.(\hat{V}_{oferta}); c\hat{V}_{oferta} + t_{(1-\alpha/2;n-k-1)}c \cdot s.e.(\hat{V}_{oferta})]$$

Colocando-se c em evidência, obtem-se:

$$IC_{venda} = [c(\hat{V}_{oferta} - t_{(1-\alpha/2;n-k-1)}s.e.(\hat{V}_{oferta})); c(\hat{V}_{oferta} + t_{(1-\alpha/2;n-k-1)}s.e.(\hat{V}_{oferta}))]$$

Em suma:

$$IC_{venda} = c \cdot IC_{oferta} \tag{11}$$

3 Estudos de Casos

3.1 EC1

3.1.1 Dados

Tabela 1:

Statistic	valor	area_total	quartos	suites	garagens	dist_b_mar	padrao
N	50	53	53	53	53	53	53
Mean	953,800.000	188.122	2.679	1.189	1.698	528.792	2.321
St. Dev.	627,318.800	116.215	0.754	0.900	0.972	308.098	0.754
Min	195,000.000	48	1	0	0	60	1
Pctl(25)	547,750.000	109	2	1	1	260	2
Pctl(75)	1,254,000.000	220	3	1	2	730	3
Max	3,000,000.000	578	4	3	4	1,430	3

3.1.2 Variância da Variável resposta

- Valores de oferta

```
var(dados$valor, na.rm = TRUE)
```

```
## [1] 393528897959
```

- Valores de oferta homogeneizados pelo fator oferta (c = 0,9)

```
var(.9*dados$valor, na.rm = TRUE)
```

```
## [1] 318758407347
```

- Ajuste da variância ($c^2\hat{\mathbb{V}}(V_{Oferta})$)

```
.9^2*var(dados$valor, na.rm = TRUE)
```

```
## [1] 318758407347
```

3.1.3 Ajuste de modelos

- Com dados de oferta

```
fit <- lm(log(valor)~area_total + quartos + suites + garagens +
            log(dist_b_mar) + I(padrao^-1), data = dados, subset = -c(31, 39))
```

- Com dados de oferta pré-ajustados

```
dados <- within(dados, valor1 <- .9*valor)
fit1 <- lm(log(valor1)~area_total + quartos + suites + garagens +
            log(dist_b_mar) + I(padrao^-1), data = dados, subset = -c(31, 39))
```

3.1.4 Estatísticas dos modelos

Tabela 2: Comparação dos modelos

	Dependent variable:	
	$log(Valor_{oferta})$	$log(Valor_{Hom})$
	(1)	(2)
Constant	13,56 (13,27, 13,86) t = 58,85***	13,46 (13,16, 13,75) t = 58,39***
area_total	0,001 (0,001, 0,002) t = 5,11***	0,001 (0,001, 0,002) t = 5,11***
quartos	0,16 (0,12, 0,21) t = 4,63***	0,16 (0,12, 0,21) t = 4,63***
suites	0,06 (0,02, 0,10) t = 1,81***	0,06 (0,02, 0,10) t = 1,81***
garagens	0,21 (0,17, 0,25) t = 6,25***	0,21 (0,17, 0,25) t = 6,25***
log(dist_b_mar)	−0,14 (−0,18, −0,11) t = −5,17***	−0,14 (−0,18, −0,11) t = −5,17***
I(padrao^-1)	−0,56 (−0,70, −0,43) t = −5,36***	−0,56 (−0,70, −0,43) t = −5,36***
Observations	48	48
R^2	0,96	0,96
Adjusted R^2	0,95	0,95
Residual Std. Error (df = 41)	0,14	0,14
F Statistic (df = 6; 41)	148,92***	148,92***
Notas:		*p<0,3; **p<0,2; ***p<0,1

3.1.5 Previsões com os modelos

a. Com dados de oferta

```
new <- dados[52, ]
p <- predict(fit, newdata = new, interval = "confidence", level = .80)
(P <- exp(p))
```

```
##         fit      lwr     upr
## 52 961660.6 924768.1 1000025
```

```
##      C.A.I.  C.A.S.
## 52 817411.5 1105910
```

- Aplicando-se o fator oferta apenas à estimativa de tendência central e deslocando-se os limites do intervalo com a mesma magnitude do valor arbitrado, conforme a NBR 14.653-02 (2011, item A.10.1.2):

```
P1 <- P
P1[, "fit"] <- 0.9*P[, "fit"]
P1[, "lwr"] <- P1[, "lwr"] - (P[, "fit"] - P1[, "fit"])
P1[, "upr"] <- P1[, "upr"] - (P[, "fit"] - P1[, "fit"])
P1
```

```
##         fit      lwr      upr
## 52 865494.6 828602.1 903858.9
```

- Aplicando-se o fator oferta a todos os parâmetros:

```
(.9*P)
```

```
##         fit      lwr      upr
## 52 865494.6 832291.3 900022.4
```

b. Com dados de oferta pré-ajustados

```
p1 <- predict(fit1, newdata = new, interval = "confidence", level = .80)
(P1 <- exp(p1))
```

```
##         fit      lwr      upr
## 52 865494.6 832291.3 900022.4
```

3.1.6 Com dados simulados

```
P1 <- P
P1[, "fit"] <- 0.9*P[, "fit"]
P1[, "lwr"] <- P1[, "lwr"] - (P[, "fit"] - P1[, "fit"])
P1[, "upr"] <- P1[, "upr"] - (P[, "fit"] - P1[, "fit"])
brf(P1)
```

```
##      fit            lwr            upr
## [1,] "  900.000,00" "  750.000,00" "1.050.000,00"
```

```
brf(.9*P)
```

```
##      fit            lwr            upr
## [1,] "  900.000,00" "  765.000,00" "1.035.000,00"
```

Como pode-se perceber, a diferença entre os intervalos obtidos passa a ser mais significante, ainda que o modelo apresente grau I de precisão.

No entanto, a situação ainda se agrava se considera-se que o limite inferior deve ser restringido pelo Campo de Arbítrio do avaliador.

```
##      fit            lwr            upr
## [1,] "  900.000,00" "  850.000,00" "1.050.000,00"
```

3.2 EC2

Foram simulados 100 dados aleatórios de venda ou oferta de imóveis segundo a expressão abaixo, em que *Fonte* é uma variável dicotômica que assume o valor zero em caso de se tratar de um dado de oferta e o valor

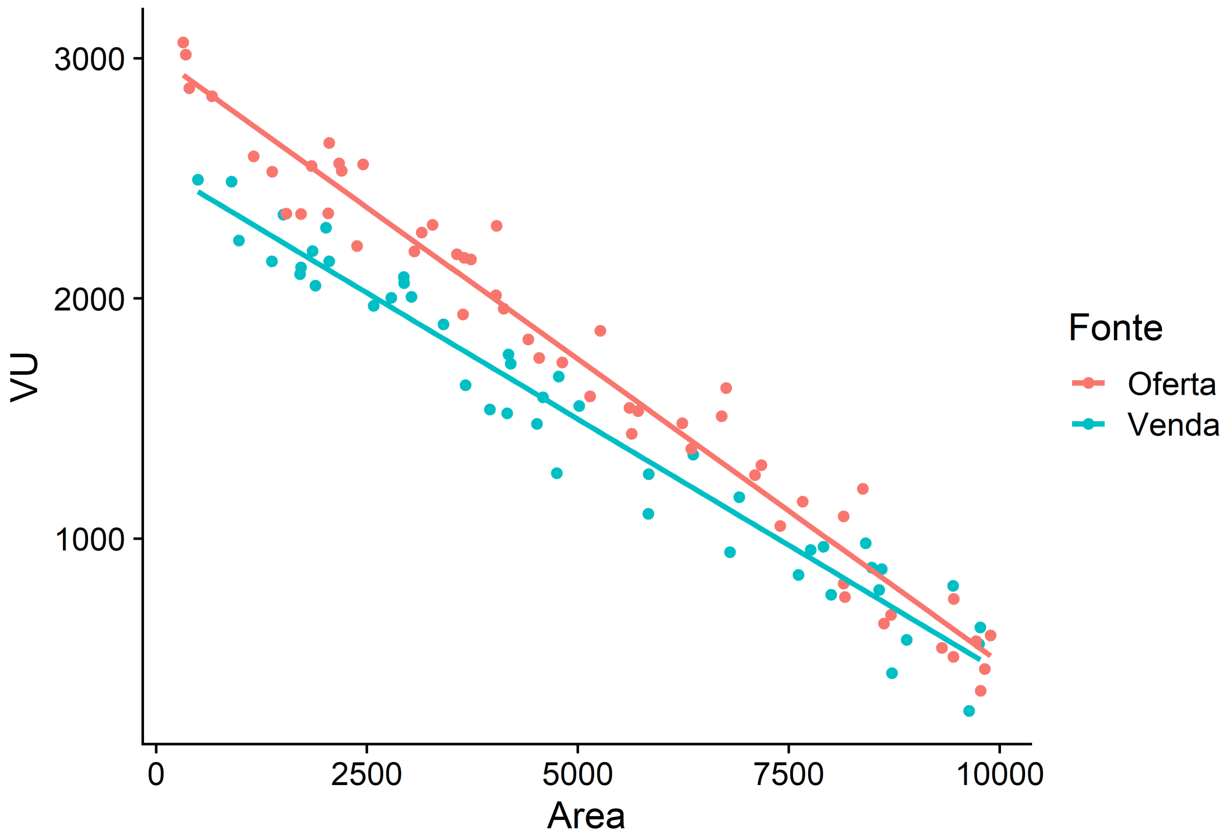


Figura 1: Dados randômicos de oferta (em vermelho) e transação (em verde)

1 no caso de se tratar de um dado de venda. Ou seja, os dados de venda foram simulados para apresentarem, em média, valor 15% menor do que os dados de oferta de imóveis. A variável $Area$ expressa a área do imóvel em metros quadrados e, por fim, a variável ε é uma variável aleatória com distribuição normal, média zero e desvio padrão igual a 150.

$$VU = (1 - 0,15.Fonte)(3000 - 0,25.Area + \varepsilon)$$

Na Figura **??** é possível visualizar os dados gerados, assim como as retas de regressão para cada tipo de dados separadamente. Como era de se esperar, os dados de maior valor unitário, sejam de oferta ou de transação, apresentam uma maior discrepância de valores entre si relativamente aos dados de maior valor unitário, à direita do gráfico.

```
## `geom_smooth()` using formula 'y ~ x'
```

Como é possível perceber já pela simples visualização gráfica dos dados, um modelo de regressão que contemple o fator fonte ou fator oferta na forma direta[1], deverá fazê-lo por meio da adição de uma interação entre a variável $Fonte$ com as outras variáveis explicativas do modelo.

Apenas para efeito ilustrativo, ajustou-se um modelo com a variável $Fonte$ na forma direta, sem a adição da interação entre ela e a variável explicativa. O modelo mostrou-se bem ajustado.

[1]Isto é, sem a utilização de transformação para a variável dependente.

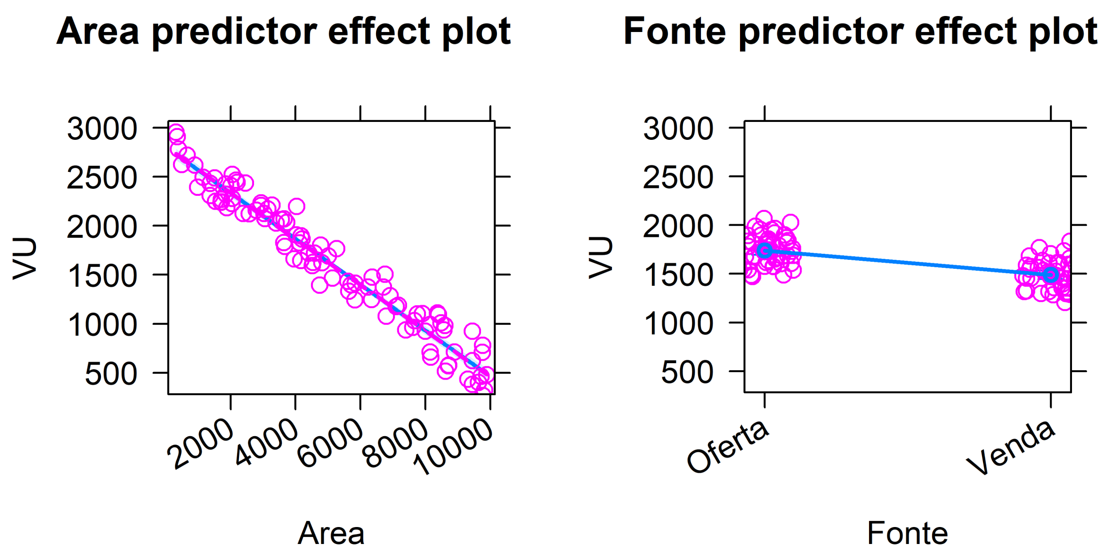


Figura 2: Fator oferta na forma direta. Má especificação do modelo.

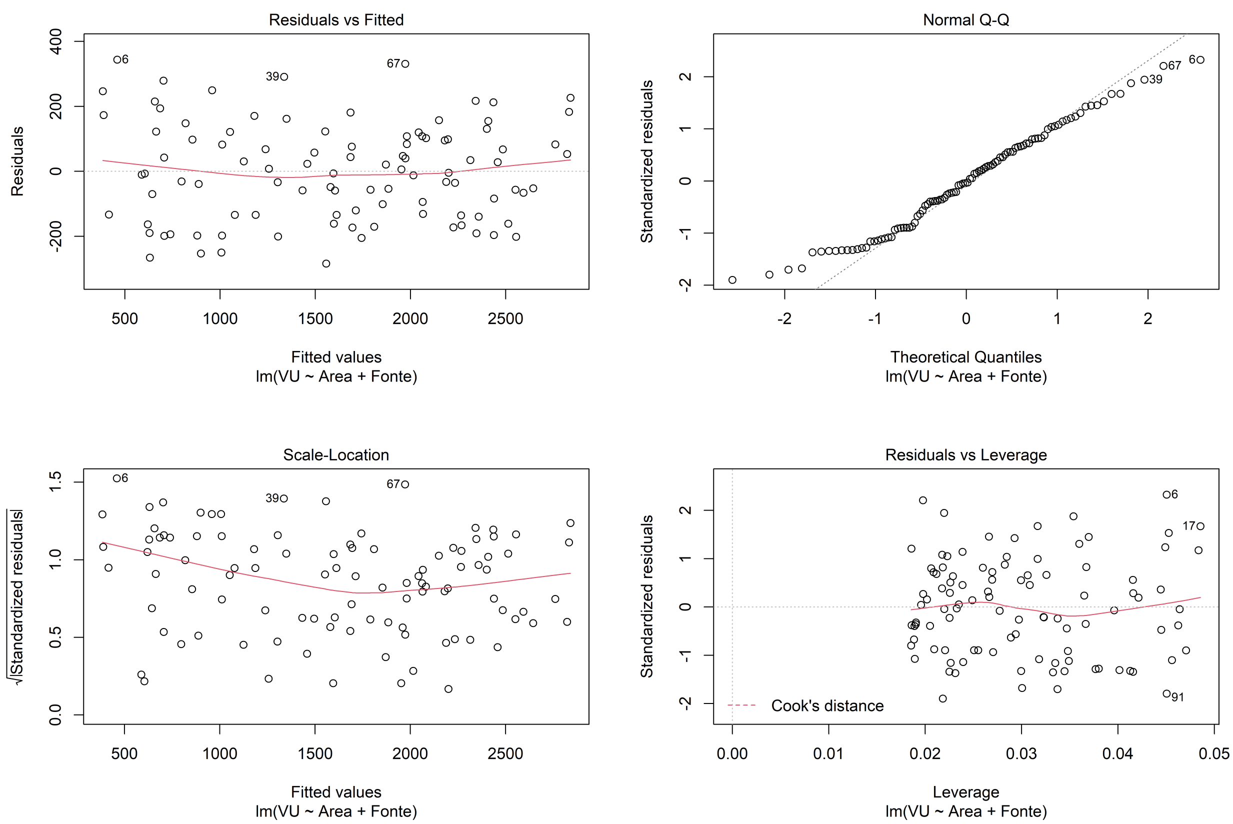


Figura 3: Diagnóstico do modelo mau especificado na forma direta.

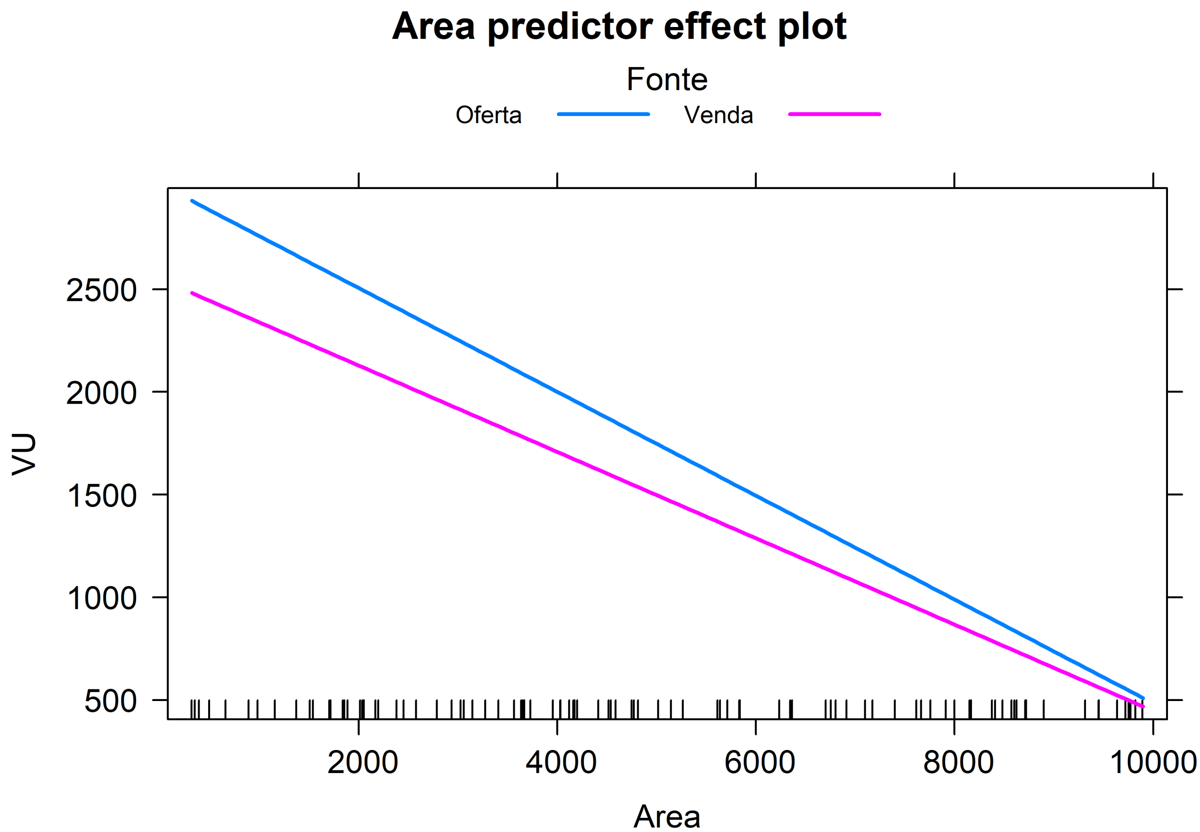
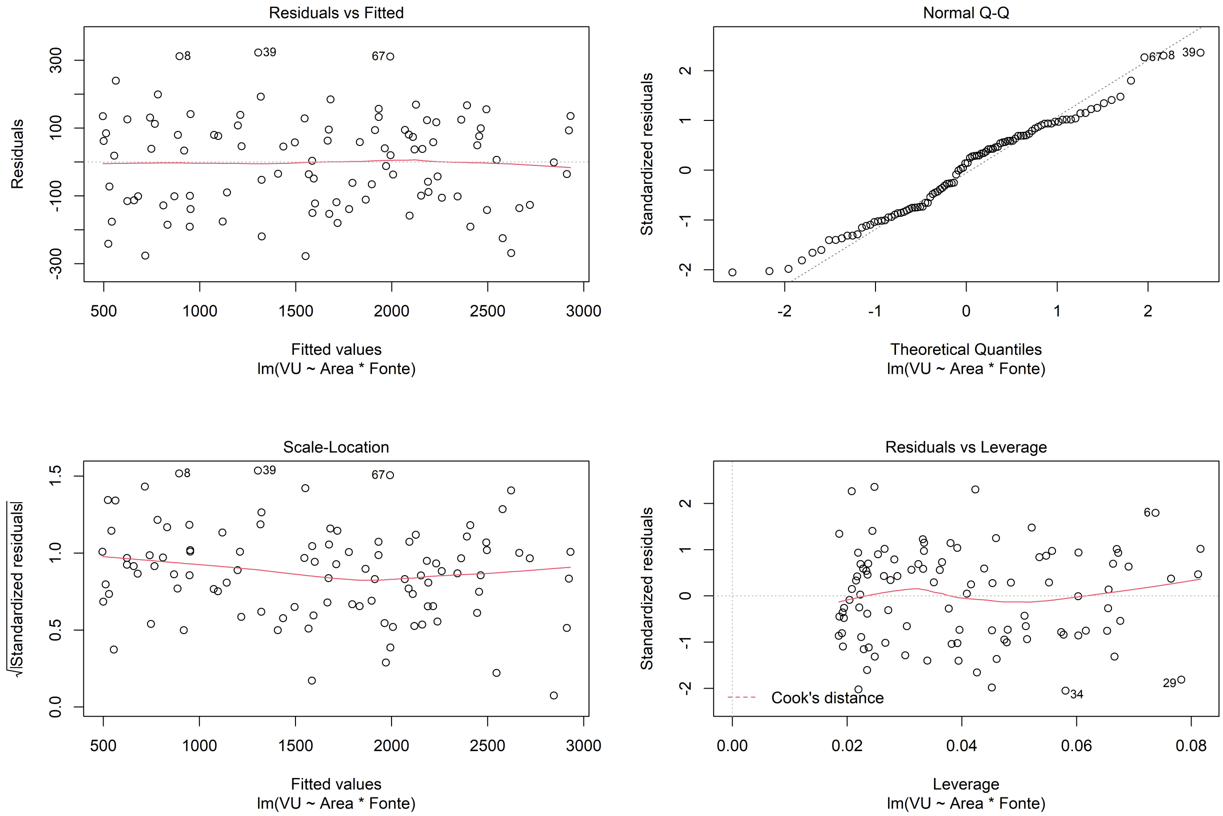


Figura 4: Fator oferta na forma direta. Modelo especificado corretamente.

Com esta formulação, o desconto é constante ao longo de todo o modelo. É possível verificar a falha na especificação do modelo de forma gráfica, pela análise dos gráficos diagnósticos da Figura 3, ou através do teste de Breusch-Pagan.

```
##
##  studentized Breusch-Pagan test
##
## data:  fit
## BP = 6.4355, df = 2, p-value = 0.04005
```

Com o ajuste correto do modelo, com a adição da interação entre a variável *Fonte* e a variável explicativa *Area*, a heteroscedasticidade desaparece.




```
##
##  studentized Breusch-Pagan test
##
## data:  fit1
## BP = 3.9476, df = 3, p-value = 0.2672
```

Uma opção dada pela ABNT (2011) é a prévia homogeneização dos dados, conforme item 9.2.1.3. A Figura **??** mostra como ficam os dados após a homogeneização.

```
## `geom_smooth()` using formula 'y ~ x'
```

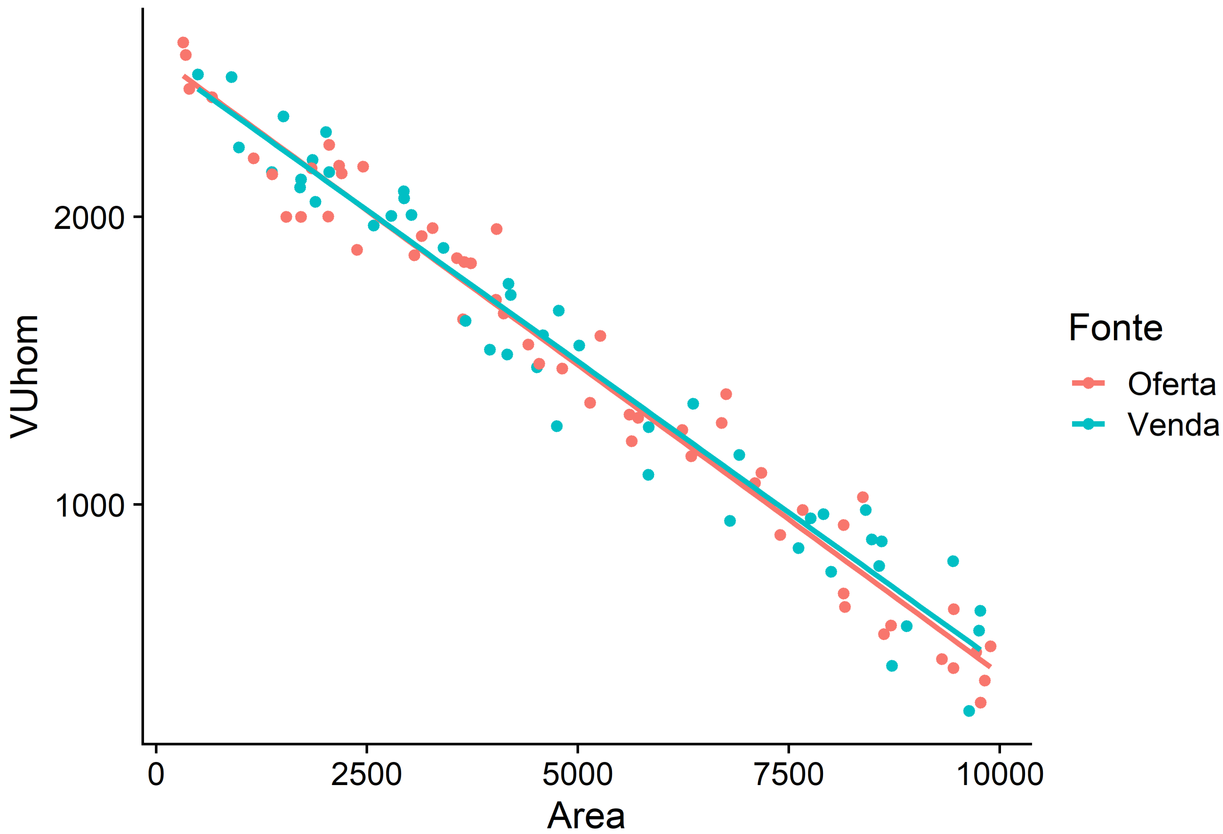
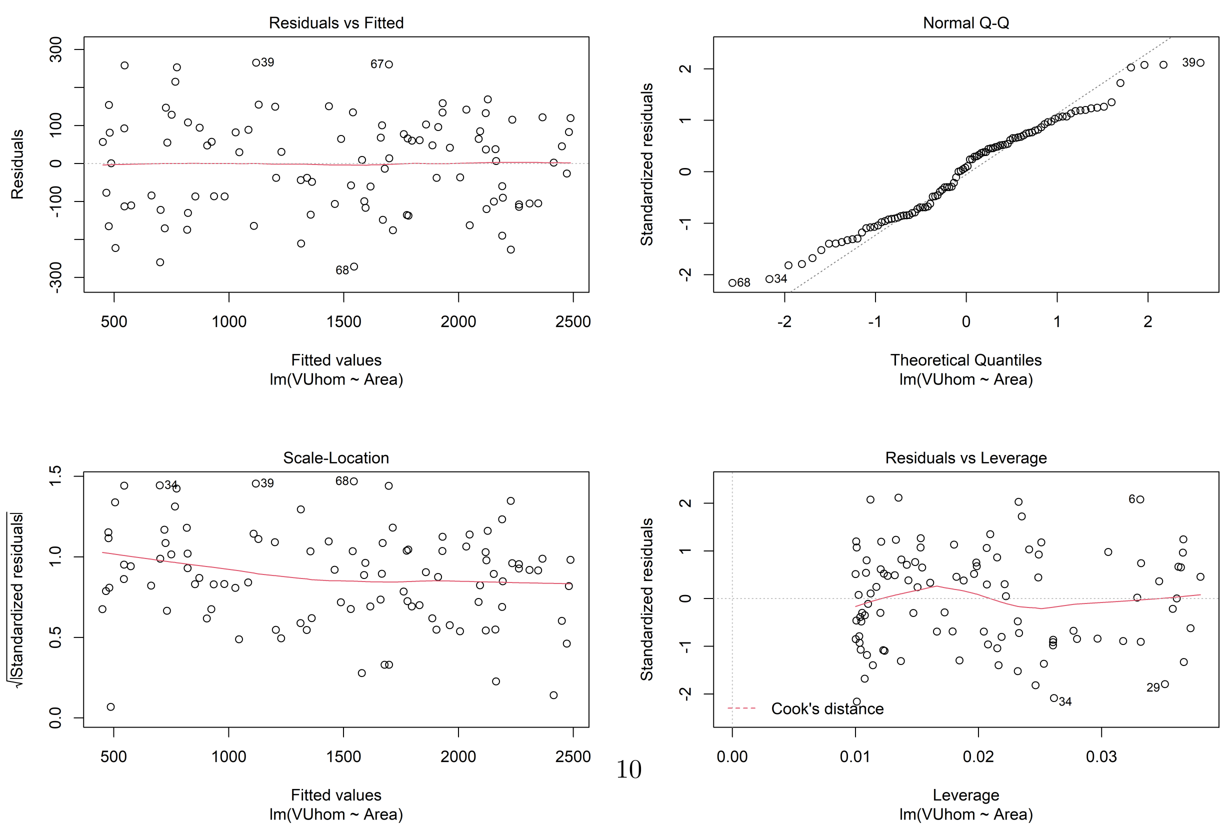


Figura 5: Dados homogeneizados.



```
##        1
## 2550.059

##        1
## 2443.899

##        1
## 2448.549

##        1
## 1498.661

##        1
## 1498.004

##        1
## 1491.235

##        1
## 330.4401

##        1
## 447.0103

##        1
## 427.5531
```

Tabela 3: Comparação dos modelos

	Dependent variable:		
	VU		VUhom
	(1)	(2)	(3)
Constant	2.914,83 (2.871,91, 2.957,75) t = 87,04***	3.011,07 (2.962,81, 3.059,33) t = 79,96***	2.554,92 (2.522,42, 2.587,41) t = 100,76***
Area	−0,23 (−0,24, −0,23) t = −44,72***	−0,25 (−0,26, −0,24) t = −39,20***	−0,21 (−0,22, −0,21) t = −48,79***
FonteVenda	−247,95 (−286,85, −209,04) t = −8,17***	−462,07 (−533,77, −390,38) t = −8,26***	
Area:FonteVenda		0,04 (0,03, 0,05) t = 4,41***	
Observations	100	100	100
R^2	0,96	0,96	0,96
Adjusted R^2	0,95	0,96	0,96
Residual Std. Error	151,29 (df = 97)	138,67 (df = 96)	126,27 (df = 98)
F Statistic	1.031,74*** (df = 2; 97)	825,26*** (df = 3; 96)	2.380,72*** (df = 1; 98)
Notas:			*p<0,3; **p<0,2; ***p<0,1

3.3 EC3

Outra possibilidade é o ajuste do modelo com a transformação da variável dependente para a escala logarítmica, o que é muito comum na Engenharia de Avaliações. A transformação, porém, não deve ser feita em função apenas da modelagem do fator oferta: uma vez averiguada a necessidade de transformação da variável dependente para a escala logarítmica, pode-se simplesmente adicionar a variável *Fonte* ao lado direto da equação de regressão, sem a necessidade da adição de termos de interação[2].

[2] Isto decorre do fato que a equação de regressão na forma logarítmica, a equação de estimação, obtida pela retransformação da equação de regressão para a forma direta, através utilização da função exponencial, relaciona os termos de uma maneira multiplicativa. Por exemplo: se a regressão for ajustada com a forma $log(Y) \sim \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, então a equação de estimação será $Y = \exp(\beta_1 X_1 + \beta_2 X_2) = \exp(\beta_1 X_1) \cdot \exp(\beta_2 X_2)$.

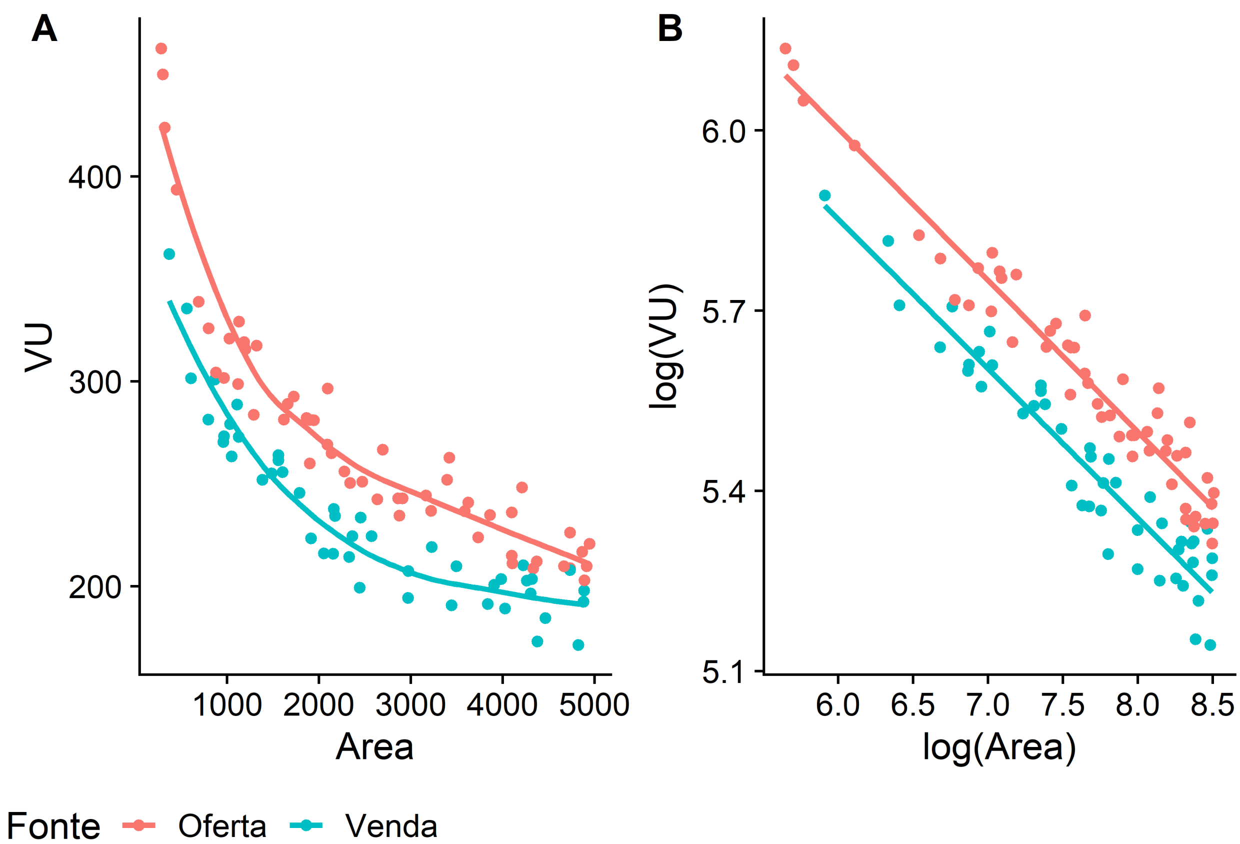


Figura 6: Dados com distribuição lognormal.

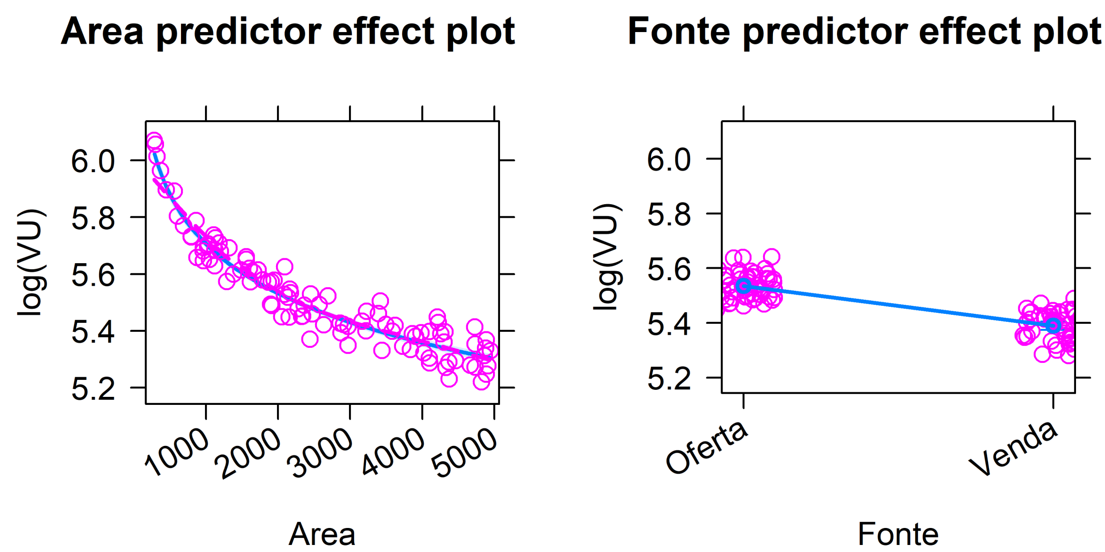
Foram gerados, então, 100 dados de oferta/transações através da seguinte expressão, com $\varepsilon \sim N(0, .05^2)$:

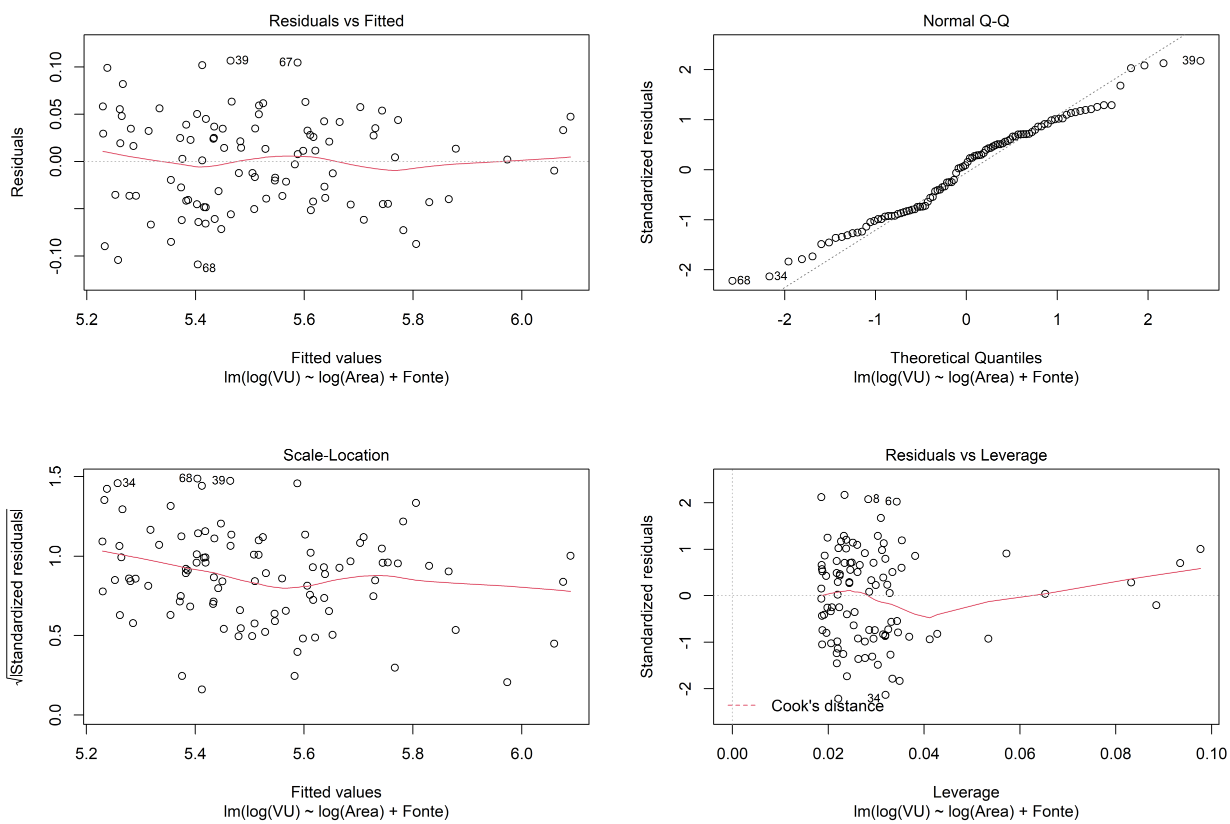
$$VU = \exp(7,5 - 0,25 \cdot \log(Area) - 0,1 \cdot Fonte + \varepsilon)$$

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```






```
##
##  studentized Breusch-Pagan test
##
## data:  fit2
## BP = 4.4404, df = 2, p-value = 0.1086
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## `geom_smooth()` using formula 'y ~ x'
```

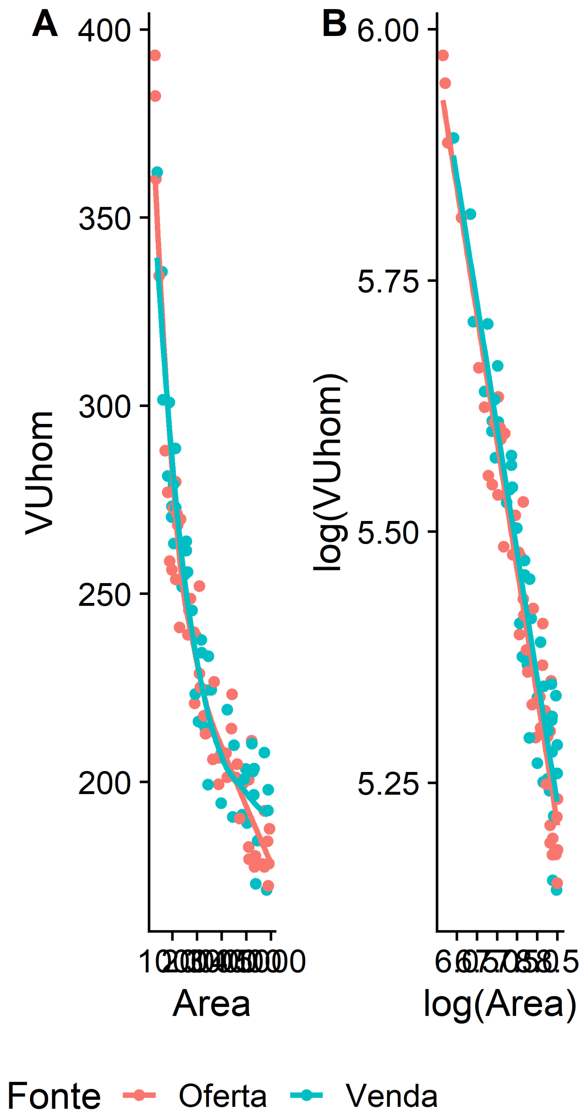


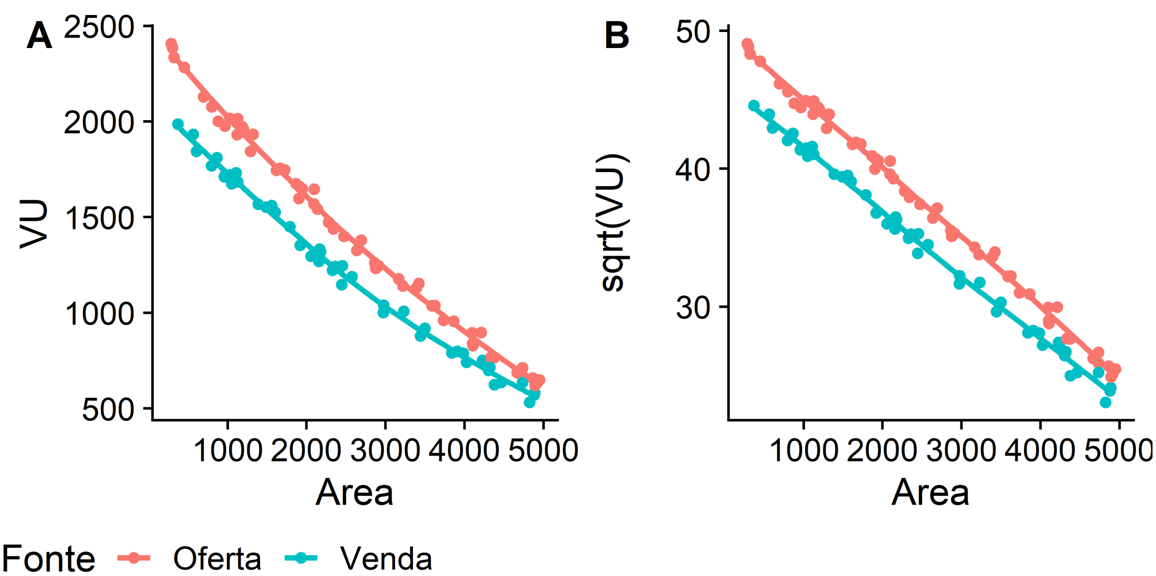
Figura 7: Dados homogeneizados.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Tabela 4: Comparação dos modelos

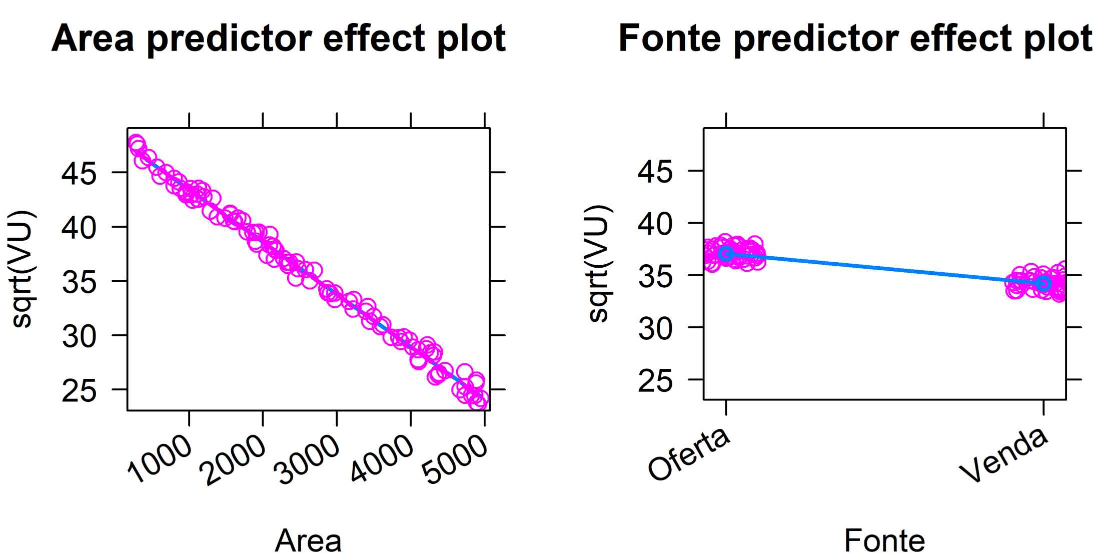
	Dependent variable:	
	log(VU)	log(VUhom)
	(1)	(2)
Constant	7,51 (7,44, 7,58)	7,35 (7,28, 7,42)
	t = 138,99***	t = 134,98***
log(Area)	−0,25 (−0,26, −0,24)	−0,25 (−0,26, −0,24)
	t = −35,80***	t = −35,40***
FonteVenda	−0,15 (−0,16, −0,13)	
	t = −14,54***	
Observations	100	100
R^2	0,94	0,93
Adjusted R^2	0,94	0,93
Residual Std. Error	0,05 (df = 97)	0,05 (df = 98)
F Statistic	756,28*** (df = 2; 97)	1.253,45*** (df = 1; 98)
Notas:		*p<0,3; **p<0,2; ***p<0,1

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
##
## Call:
## lm(formula = sqrt(VU) ~ Area + Fonte, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03545 -0.41809 -0.02231  0.41313  1.37505
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.955e+01  1.292e-01  383.47   <2e-16 ***
## Area        -4.828e-03  4.006e-05 -120.51   <2e-16 ***
```

```
## FonteVenda  -2.848e+00  1.134e-01  -25.11   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5652 on 97 degrees of freedom
## Multiple R-squared:  0.9936, Adjusted R-squared:  0.9935
## F-statistic:  7562 on 2 and 97 DF,  p-value: < 2.2e-16
```



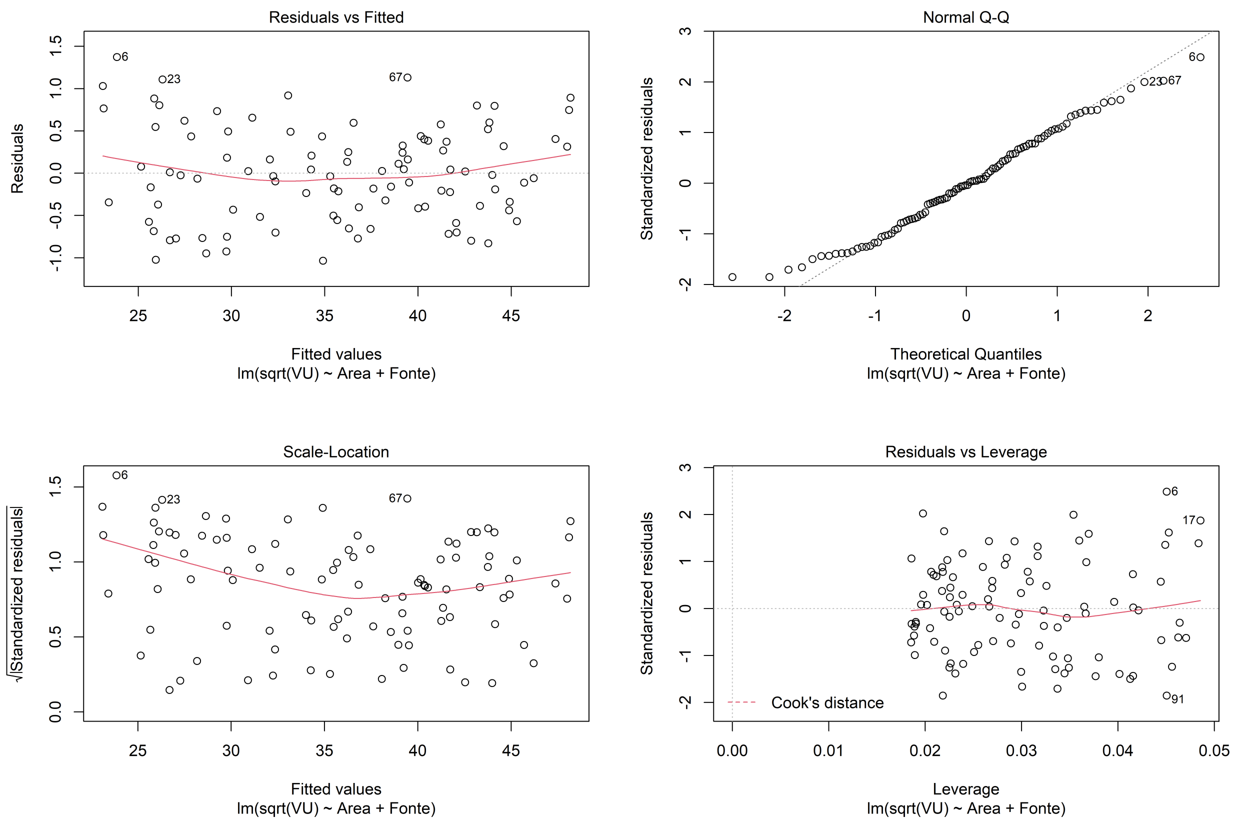


Figura 8: Modelo mau especificado. Fator oferta.

```
## ‘geom_smooth()‘ using method = ’loess’ and formula ’y ~ x’

## ‘geom_smooth()‘ using formula ’y ~ x’

## ‘geom_smooth()‘ using method = ’loess’ and formula ’y ~ x’
```

4 Conclusão

Quando da utilização de dados de oferta para elaboração de modelos, além do valor da estimativa central, também os limites do intervalo de confiança devem ser ajustados pelo fator oferta.

Pelo procedimento preconizado na atual normativa, o intervalo de confiança é simplesmente transladado. Isso implica que o valor inferior do intervalo de confiança acaba minorado por um fator superior ao fator oferta, enquanto o limite superior do intervalo de confiança é minorado com um fator menor do que o fator oferta.

Quando a amplitude do intervalo de confiança é relativamente pequena, como no estudo de caso apresentado, esta diferença tende também a ser pequena, podendo até, em alguns casos, ser negligenciável.

No entanto, caso a amplitude do intervalo de confiança seja relativamente grande, isto passa a ser um problema, pois os intervalos de valores admissíveis calculados conforme preconiza a NBR 14.653-02, tendem a tornar-se fortemente assimétricos, como ilustrado pelo segundo caso apresentado, com dados simulados, já que o limite do campo de arbítrio inferior é atingido.

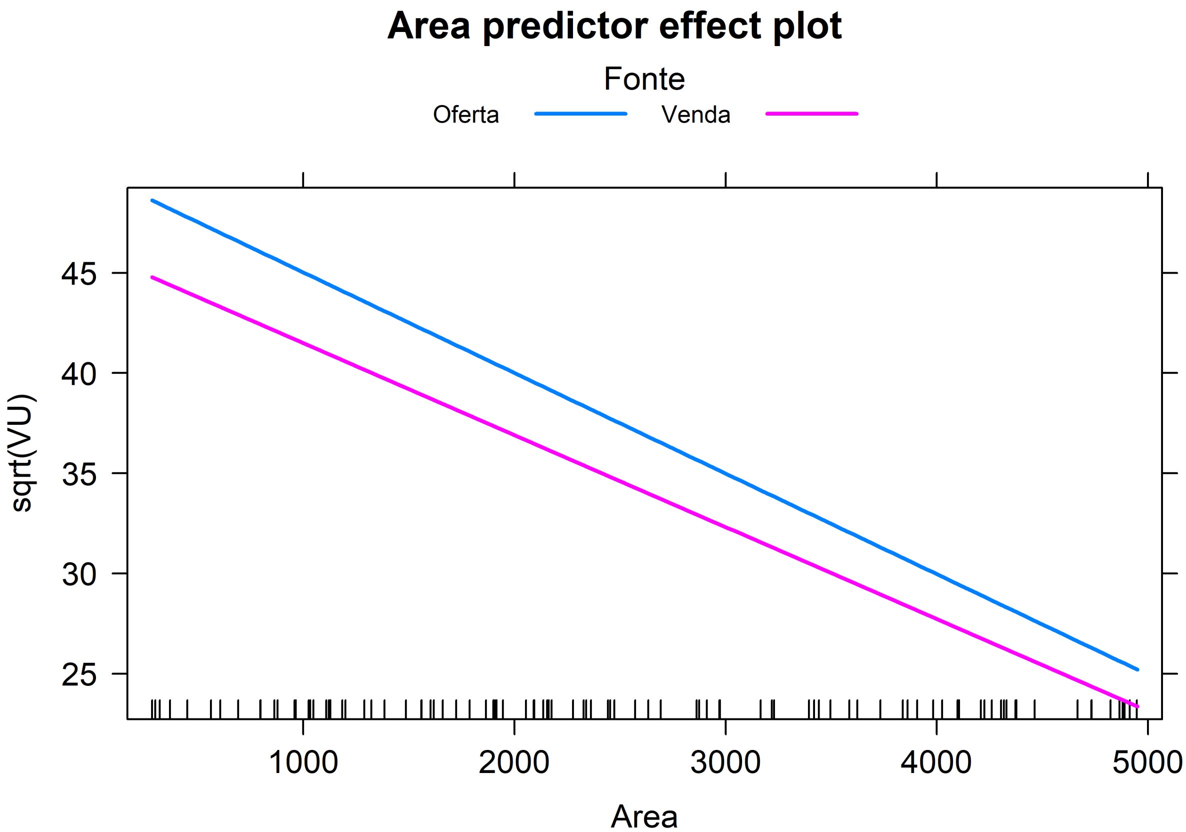


Figura 9: Fator oferta com transformação raiz quadrada. Modelo especificado corretamente.

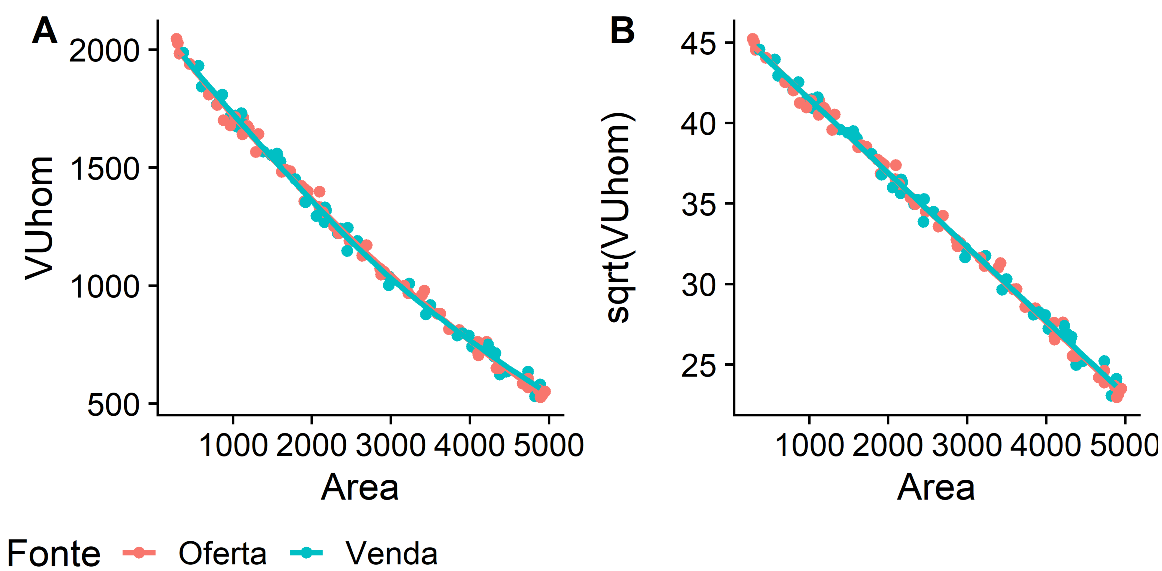


Figura 10: Dados homogeneizados. Transformação raiz quadrada.

Referências

ABNT. 2011. *NBR 14653-2: Avaliação de Bens – Parte 2: Imóveis Urbanos*. Rio de Janeiro: Associação Brasileira de Normas Técnicas.

Faraway, J. J. 2004. *Linear Models with R*. Chapman & Hall/Crc Texts in Statistical Science. Taylor & Francis. https://books.google.com.br/books?id=fvenzpofkagC.

Matloff, Norman Saul. 2009. *From Algorithms to Z-Scores: Probabilistic and Statistical Modeling in Computer Science*. Davis, California: Orange Grove Books. http://heather.cs.ucdavis.edu/~matloff/132/PLN/probstatbook/ProbStatBook.pdf.