Curso de Fatores

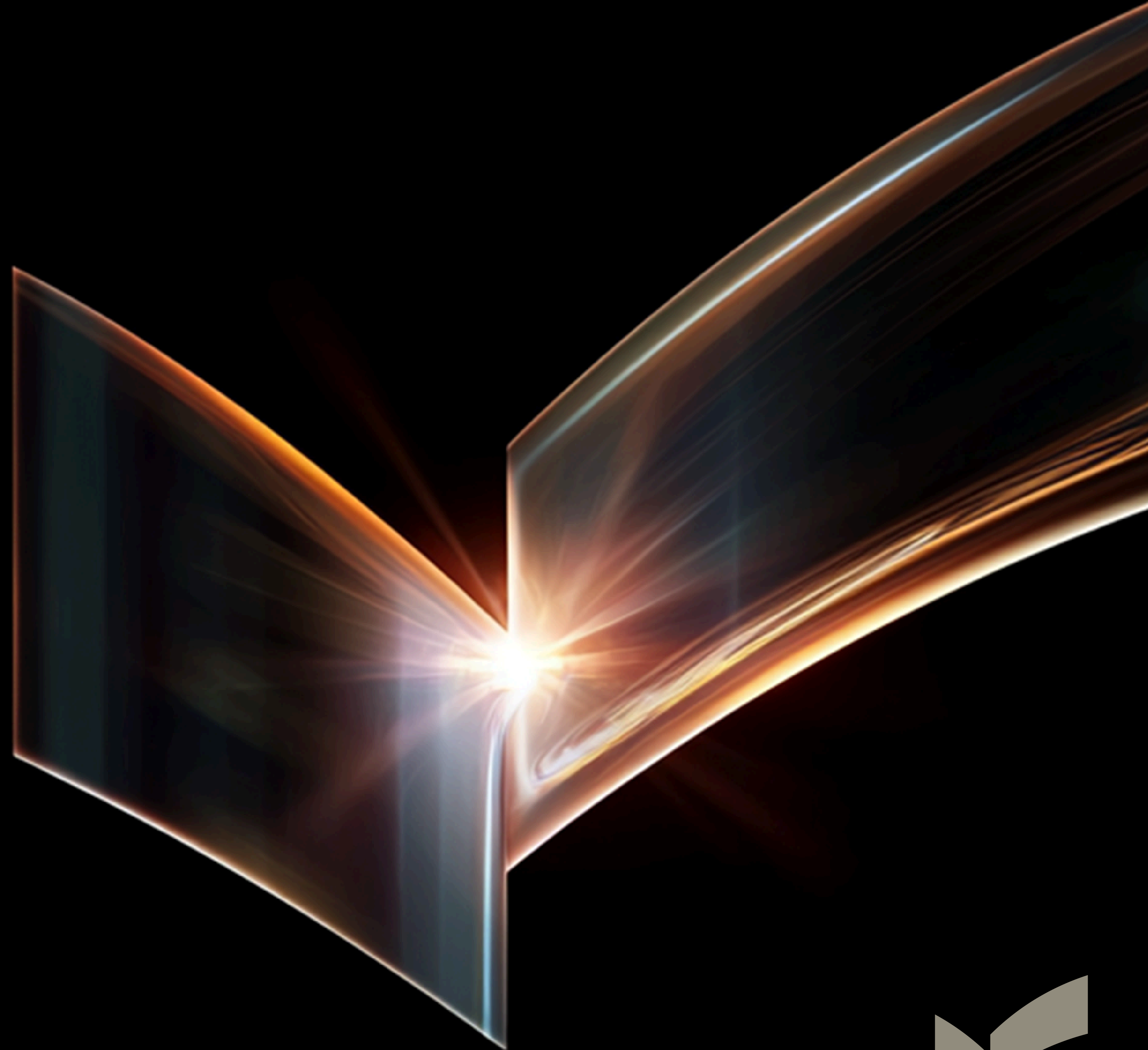
Fatores Multiplicativos Fundamentados

Luiz Droubi
Academia da Engenharia de Avaliações

25 de julho de 2025

Índice

Introdução



Sobre a arte de comunicar

Comunicação não é o que você fala, é o que o outro entende!

- O Tratamento Científico trouxe **consistência** para a Engenharia de Avaliações
- O Tratamento por Fatores, contudo, não deixou de ser utilizado
  - Acredita-se que, em parte, isso se deve à clareza obtida com este tipo de tratamento
- A consistência alcançada com o Tratamento Científico deu-se a custa de uma perda na clareza da comunicação
- É possível conciliar?
  - Trivelloni (2005);
  - Cerino et al. (2020);
  - Droubi, Zilli, e Hochheim (2021);
  - Droubi e Florencio (2024) e Oliveira, Bandeira, e Távora (2024)

Sobre a necessidade de Fatores

- Há outra utilidade, além da clareza, na utilização do tratamento por fatores?
- Na confecção de Plantas de Valores Genéricos (PVG):
    - Ao elaborar PVG por face de quadra:
    - Na fase de avaliação, ajustamos um modelo de regressão linear
    - Depois, prevemos valores para os lotes paradigma em cada face de cada quadra (*e.g.* 12x36m^2)
    - Porém, esta PVG precisa ser utilizada para prever valores dos lotes reais da cidade, que tem características que podem ser diferentes das do lote paradigma
    - Ao calcular o valor dos lotes reais, precisamos ajustar os valores de face de quadra para as características reais dos lotes
    - Ao comunicar à população o motivo das diferenças nas avaliações de cada imóvel, os fatores são uma boa pedida

Sobre a necessidade de Fatores

- Nas avaliações pontuais:
  - Ao realizar uma avaliação imobiliária, nos deparamos com variáveis incompletas, mas que deveriam ser utilizadas. Por exemplo:
    - A elasticidade das ofertas em relação aos preços, nas avaliações imobiliárias em geral, urbanas ou rurais;
    - O andar em que se situa o imóvel, na avaliação de um apartamento;
    - A pedologia do solo, ao avaliar um lote para construção/incorporação.
- Que bom seria se sempre tivessemos informações completas e claras sobre os imóveis
  - Mas então não precisaríamos de avaliadores!

Origem dos Fatores Multiplicativos

Qual a origem dos fatores multiplicativos?

- A utilização de fatores multiplicativos parece ser mais intuitiva do que a aplicação de fatores aditivos
- Ao colocar o seu imóvel à venda é natural comparar o seu imóvel a um imóvel parecido que tenha sido recentemente vendido
- Se um apartamento foi vendido no seu prédio, com características praticamente idênticas às do seu apartamento
  - Imaginemos que essa venda tenha se dado ao valor de mercado do apartamento
  - Qual a oferta mínima pelo seu apartamento você estaria disposto a aceitar?

Qual a origem dos fatores multiplicativos? (2)

- Comparando com o apartamento do seu vizinho, você faz as seguintes observações:
  - O seu apartamento situa-se numa posição mais privilegiada no edifício
    - Por exemplo, uma posição solar mais favorável
    - Ou um andar mais alto
- Um apartamento situado em uma fachada mais favorável (ou em um andar mais alto) deve ter um valor um pouco superior a um apartamento com as mesmas características, porém situado em uma fachada desfavorável (ou em um andar mais baixo).
  - O quão superior é que são elas!!!

Qual a origem dos fatores multiplicativos? (3)

- Se pensarmos em R$ 20.000,00 a mais por um apartamento de mesmas características, porém situado numa fachada mais favorável, isto é pouco ou é muito?
  - Se o valor de venda do apartamento similar foi R$ 200.000,00, então isto equivale a 10,0% a mais, o que é relevante;
  - Porém, se o valor de venda do apartamento similar foi R$ 2.000.000,00, R$ 20.000,00 é praticamente insignificante (1,0%)!
- Se pensarmos em um valor 10% maior para um apartamento de mesmas características, porém em situação mais privilegiada, temos:
  - Apartamento de R$ 200.000,00: R$ 20.000,00 a mais!
  - Apartamento de R$ 500.000,00: R$ 50.000,00 a mais!
  - Apartamento de R$ 900.000,00: R$ 90.000,00 a mais!

Qual a origem dos fatores multiplicativos? (3)

- Se olharmos pela ótica do comprador, a situação é ainda mais clara?
  - Se você está procurando um apartamento para comprar e se depara com duas oportunidades:
    - Um apartamento ao preço de R$ 200.000,00;
    - Outro apartamento no mesmo edifício, porém em posição mais favorável, por R$ 220.000,00;
      - A sua escolha irá depender da sua preferência!
        - Ambos os apartamentos estão corretamente precificados

Qual a origem dos fatores multiplicativos? (4)

- Porém, se você se depara com duas ofertas:
  - Um apartamento ao preço anunciado de R$ 2.000.000,00;
  - Outro ap. no mesmo edifício, em posição mais favorável, por R$ 2.020.000,00;
  - Obviamente você irá optar pelo segundo, pois a diferença de preço é mínima e, por apenas 1% a mais você poderá desfrutar de um apartamento melhor
    - Os apartamentos não estão bem precificados
- Numa negociação, é comum a situação:
  - Oferta de R$ 1,8 mi pelo ap. situado em posição solar desfavorável;
  - E justificar a proposta assim:
    - "Tem um apartamento ao lado, com fachada norte, por R$ 2.020.000,00."
  - A questão é que o preço do melhor é que pode estar com desconto
    - Por isso precisamos de avaliadores!

Qual a origem dos fatores multiplicativos? (5)

- Se eu estou certo, um *Fator posição solar* deverá se apresentar na forma:
  - $$F_{PS} = \frac{PS_i}{PS_{paradigma}} = \begin{cases} 1,0 & \text{se } PS_i = \text{Neutra} \\ 1,1 & \text{se } PS_i = \text{Favorável} \\ 0,9 & \text{se } PS_i = \text{Desfavorável} \end{cases}$$
- Se o valor de mercado de um apartamento numa fachada neutra é R$ 1.000.000,00
  - Então o valor de um apartamento similar, porém em fachada favorável será:
    - $$P_i = \overline{P}_{Hom} \cdot F_{PS} = 1.000.000,00 \cdot 1,1 = \text{R\$}1.100.000,00$$

Utilização de Fatores Multiplicativos

Como utilizar fatores multiplicativos

- Para homogeneizar valores (Lima 2006):
  - $$P_{Hom_i} = \frac{P_{Obs_i}}{F_{1i} \cdot F_{2i} \cdot \ldots \cdot F_{ki}} \qquad (1)$$
- Exemplo:
  - Um imóvel i da amostra conta com área e frente diferentes do avaliando
    - Primeiro precisamos homogeneizar este elemento em relação ao avaliando:
    - $$PU_{Hom_i} = \frac{PU_i}{F_{Area} \cdot F_{frente}}$$
    - Então, uma vez efetuada a homogeneização, todos os valores homogeneizados serão utilizados para o cômputo de um $\overline{PU}_{Hom}$, que será utilizado para estimar o valor de mercado do avaliando

Como utilizar fatores multiplicativos

- Para prever valores (Lima 2006):
  - $\hat{P}_i = \overline{P}_{Hom} \cdot F_{1i} \cdot F_{2i} \cdot \ldots \cdot F_{ki} \qquad (2)$

Fator Área de Abuhnaman

- Fator área de Abuhnaman (adaptado):

$$F_A = \left(\frac{A_i}{A_{paradigma}}\right)^{-1/4}$$

- Exemplo:
  - $A_{paradigma} = 360m^2$
  - $A_i = 400m^2$
  - $P_i = \text{R\$ } 500.000,00$
  - $F_{Ai} = \left(\frac{400}{360}\right)^{-1/4} \approx 0,9306$
  - $P_{Hom_i} = P_i / F_{Ai} = 500.000,00/0,9306 = \text{R\$ } 537.285$
    - Se a área do imóvel fosse igual a 360m^2, então seu preço seria maior!
- Também é um fator multiplicativo!
  - De onde vem isto?

Qual a origem dos fatores multiplicativos?

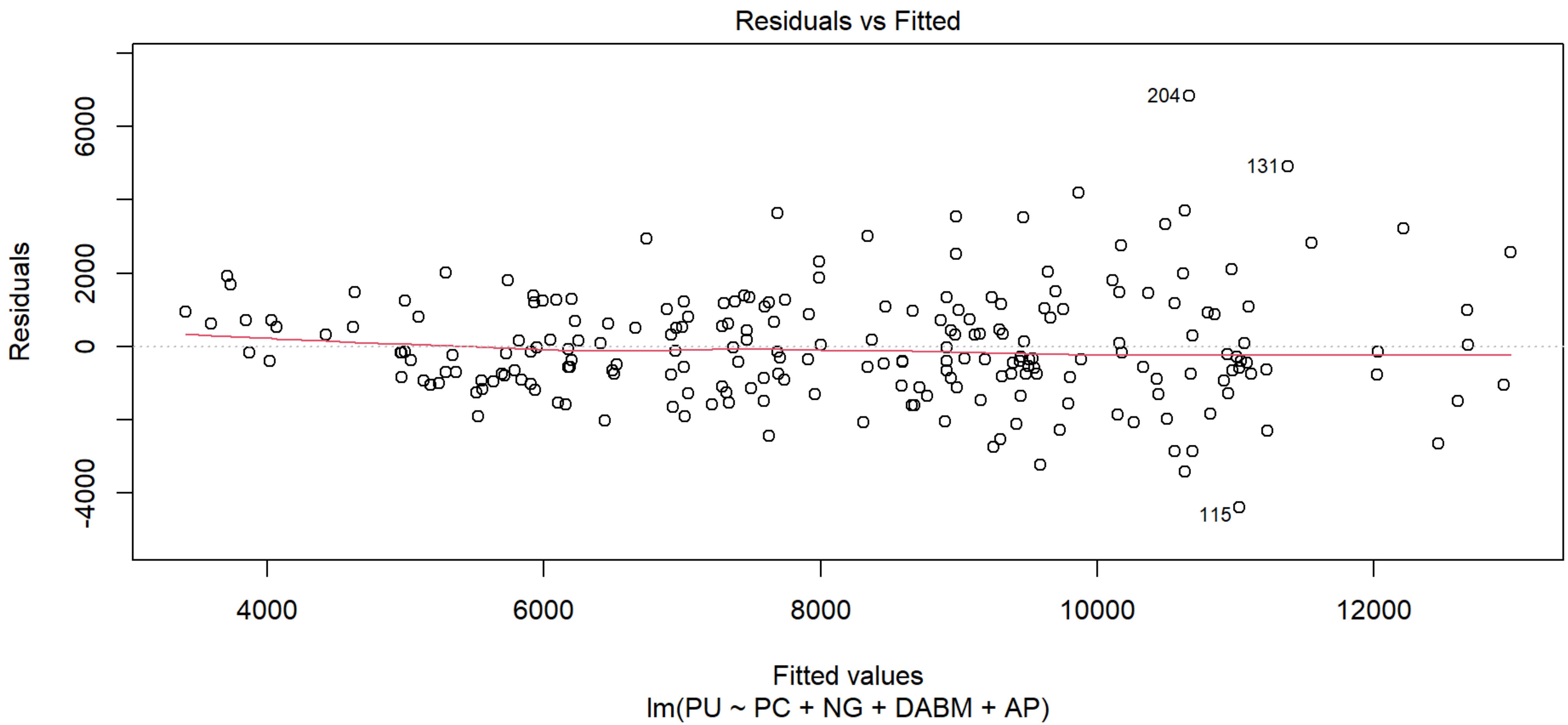


Figura 1: Resíduos vs. valores ajustados para um modelo aditivo.

Cobb-Douglas

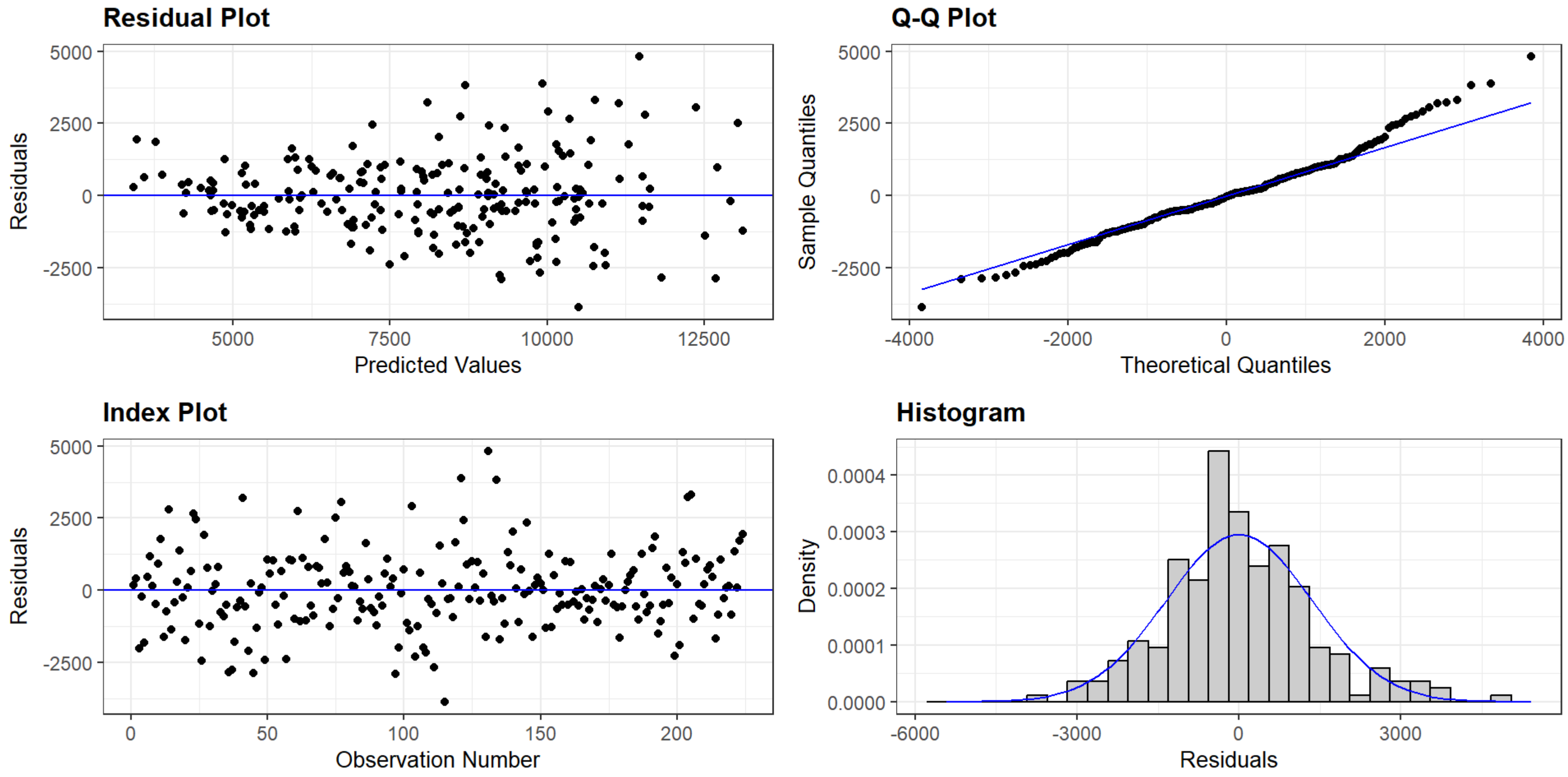
- Na econometria é muito conhecida a Função de *Cobb-Douglas*
  - Na sua forma mais simples:
    - $$Y = a.\, X^b$$
- É possível linearizar a função de *Cobb-Douglas*:
  - $$\ln(Y) = \ln(a) + b.\ln(X)$$
- O que torna fácil estimar a regressão:
  - $$\ln(Y) = \beta_0 + \beta_1.\ln(X) + \varepsilon$$
- Uma vez estimado o modelo de regressão acima, pode-se obter $\hat{a}$ e $\hat{b}$:
- $\hat{a} = \exp(\hat{\beta}_0);\ \hat{b} = \hat{\beta}_1$
- $Y = \hat{a}X^{\hat{b}}$

Cobb-Douglas (2)

- A linearização está centrada na hipótese de que o erro ξ é **multiplicativo**:
  - $$Y = \hat{a}X^{\hat{b}} \cdot \xi$$
  - $$\xi = \exp(\hat{\varepsilon}) = \frac{Y}{\hat{a}X^{\hat{b}}} = \frac{Y}{\hat{Y}}$$
- Explicação:
  - $$\begin{aligned} \ln(Y) &= \ln(\hat{a}) + b \cdot \ln(X) + \hat{\epsilon} \\ \exp[\ln(Y)] &= \exp[\ln(a) + b \cdot \ln(X) + \hat{\epsilon}] \\ Y &= \exp[\ln(\hat{a})] \cdot \exp[b \cdot \ln(X)] \cdot \exp(\hat{\epsilon}) \\ Y &= \hat{a} \cdot X^{\hat{b}} \cdot \exp(\hat{\epsilon}) \\ \exp(\hat{\epsilon}) &= \frac{Y}{\hat{a} \cdot X^{\hat{b}}} \end{aligned}$$

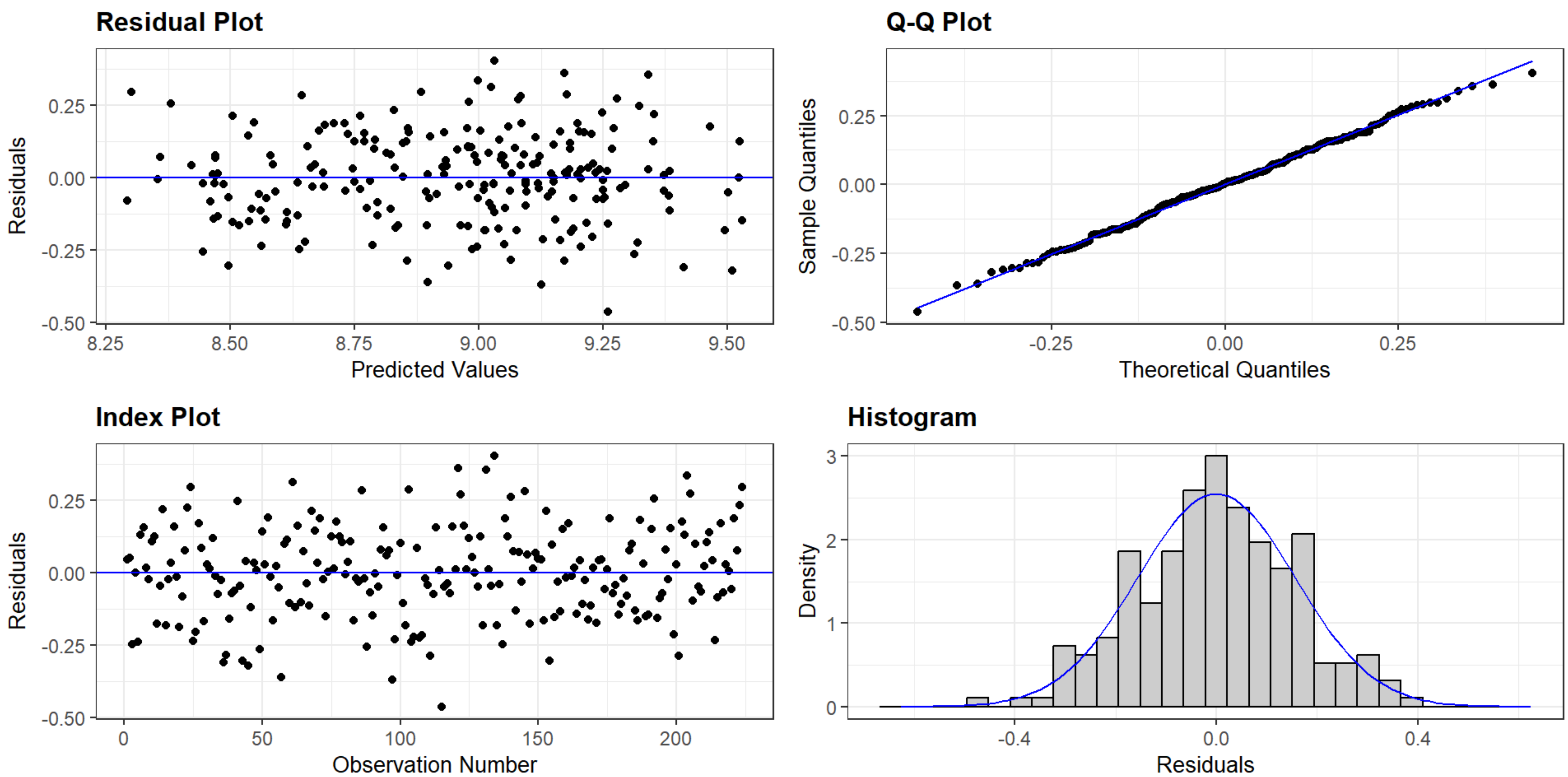
Cobb-Douglas (3)

- A hipótese de que o erro é aditivo pode não ser verificada no mercado imobiliário:



Termo de erro aditivo.

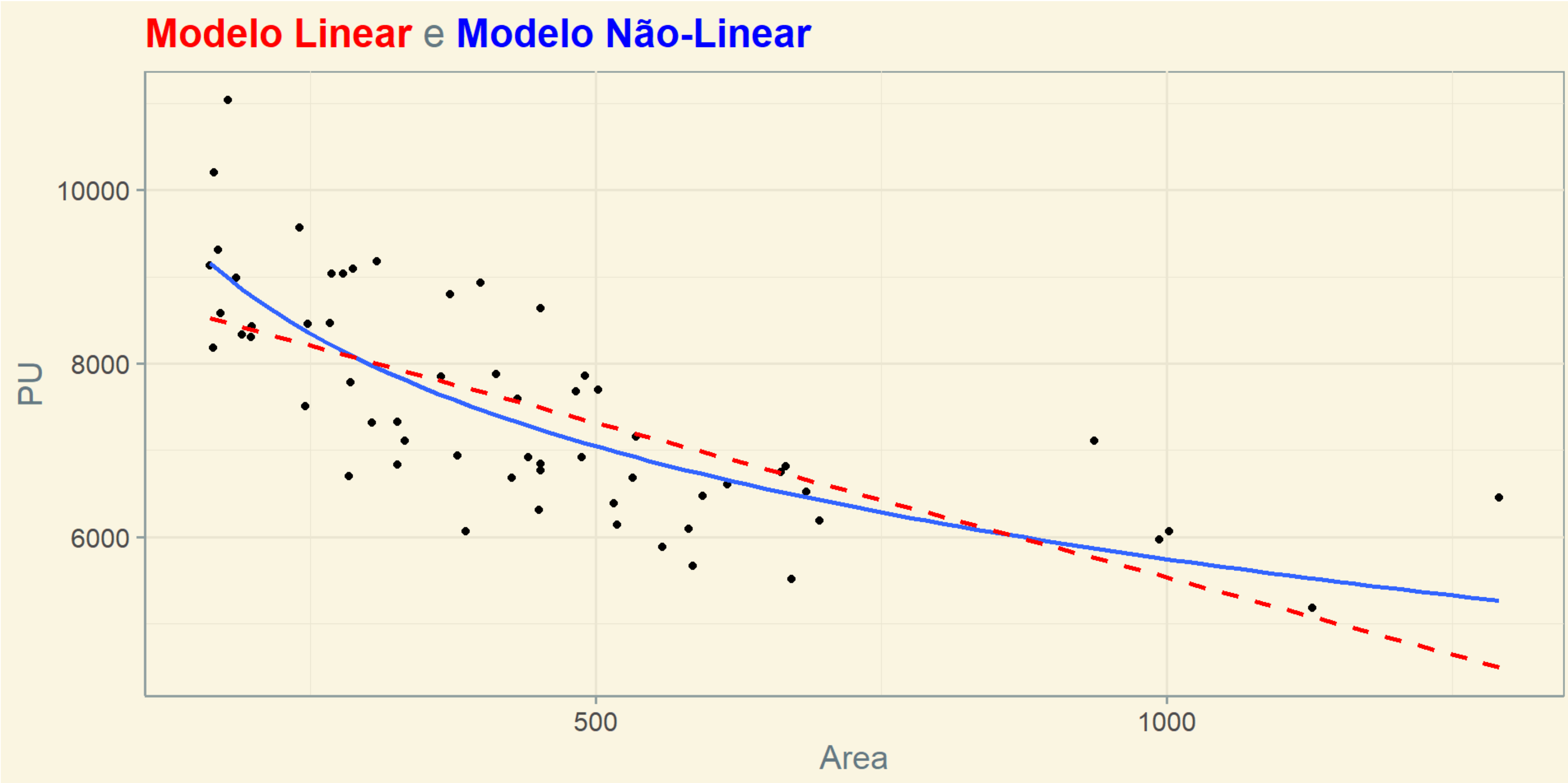
Cobb-Douglas (4)



Termo de erro multiplicativo

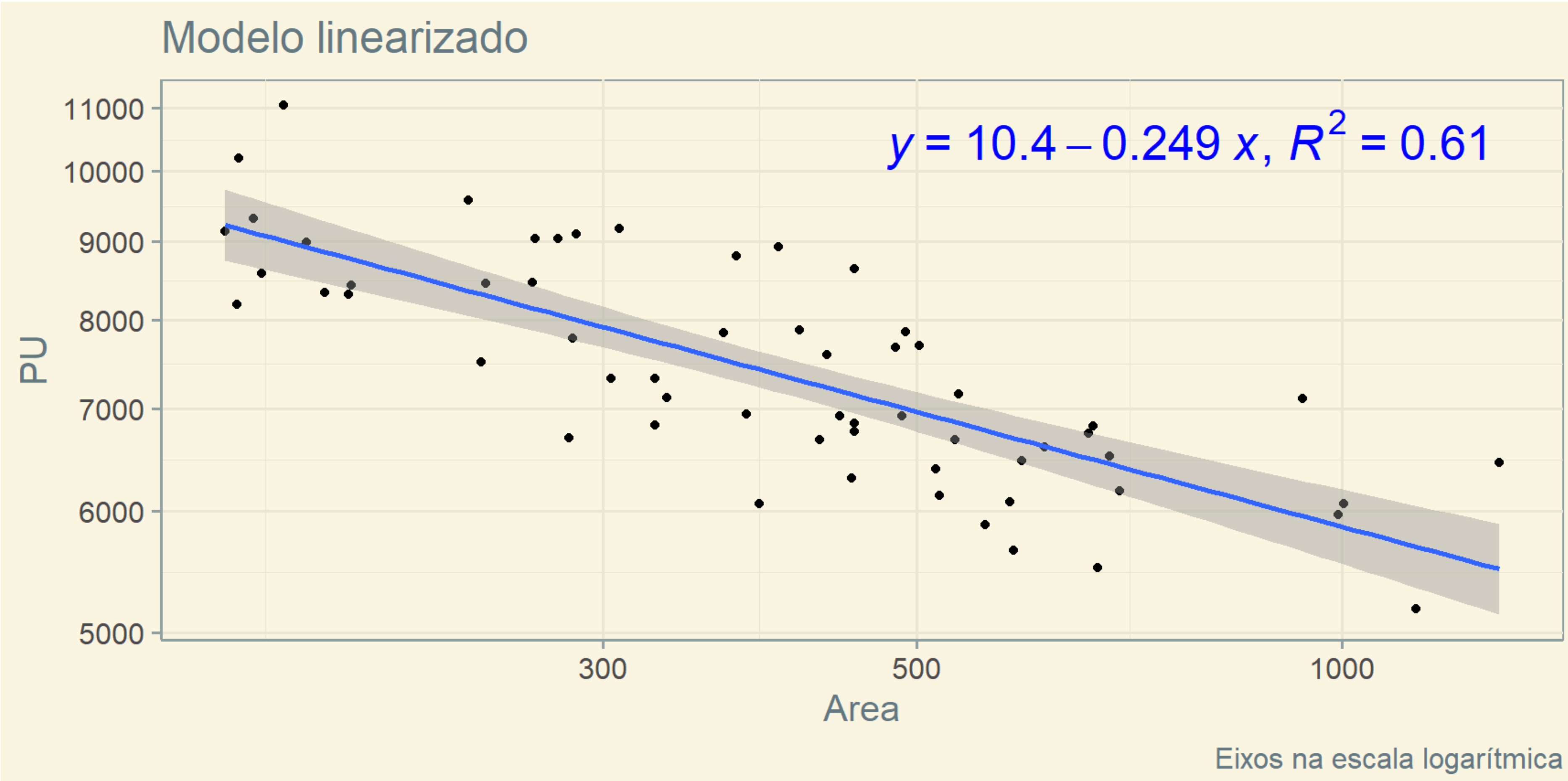
- A hipótese dos erros multiplicativos parece mais adequada!

Cobb-Douglas (5)



Dados para exemplificar a equação de Cobb-Douglas.

Cobb-Douglas (6)



- $a = \exp(10,4) \approx 32.860,00;\ b \approx -0,25$
- $PU = 32.860,00.Area^{-0,25}$

Centralização de Dados

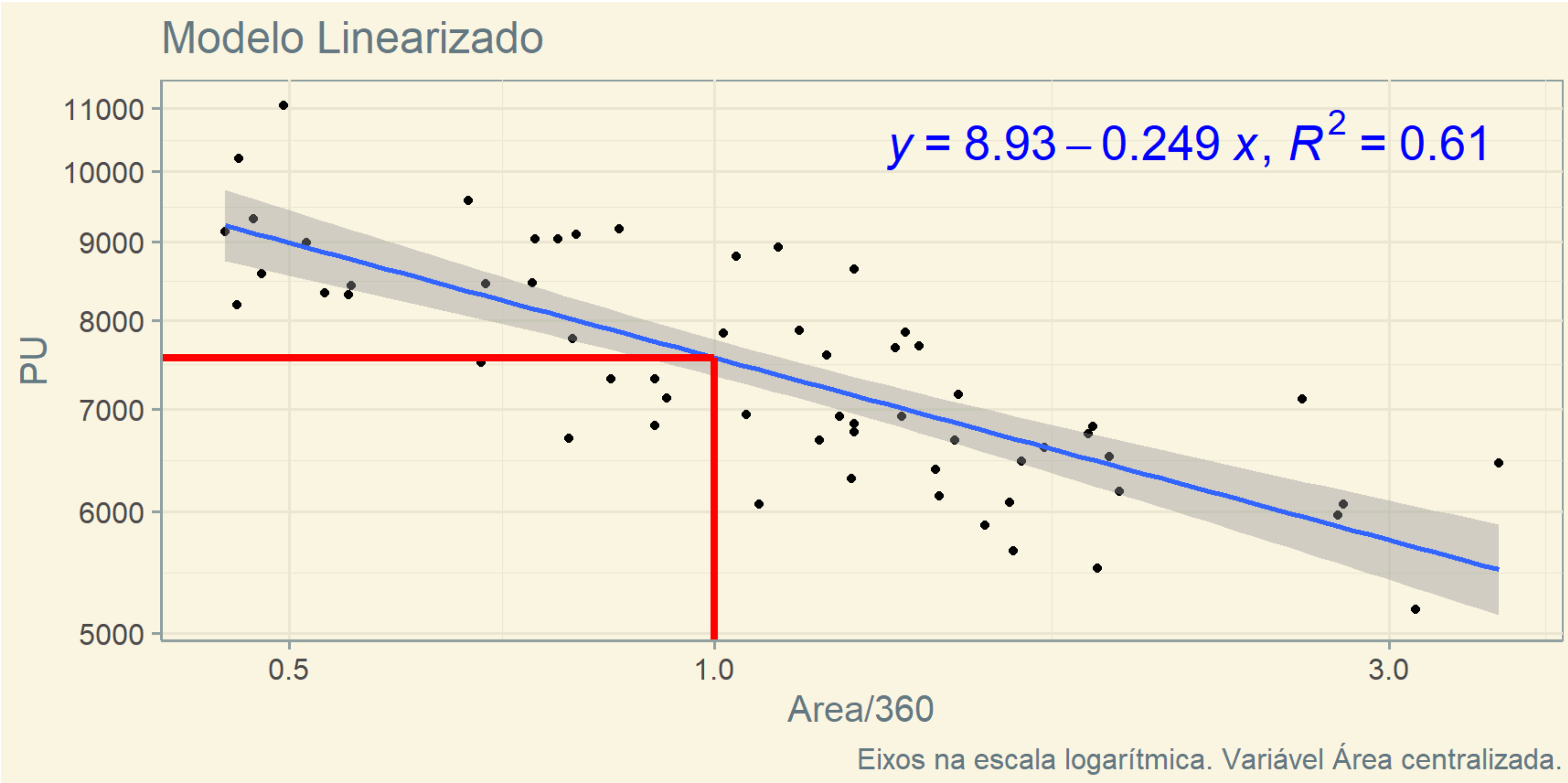
Justificativa

- O valor de $\hat{\beta}_0$ não apresenta significado físico:
  - Não existe lote com $0m^2$ de área
- A equação de estimação, portanto, fica com a interpretação um tanto prejudicada
- Isto pode ser contornado facilmente, contudo.
  - Basta centralizar a variável explicativa em relação à suas média (ou ao valor do imóvel avaliando)
    - Por exemplo, se o lote avaliando apresenta área de ($360m^2$)
    - No exemplo acima, se ao invés de utilizarmos a área do lote como variável explicativa, utilizarmos a área do lote dividida por 360, teremos:
    - $$\ln(PU) = \beta_0 + \beta_1 \cdot \ln(Area/360) + \epsilon$$

Justificativa (2)

- $$\ln(PU) = \beta_0 + \beta_1 \cdot \ln(Area/360) + \epsilon$$
    - Assim, quando a área do lote for igual a $360m^2$, então:
        - $\ln(Area/360) = \ln(360/360) = \ln(1) = 0$
        - $\widehat{PU} = \hat{\beta}_0$, ou seja, $\hat{\beta}_0$ irá representar $\widehat{PU}$ de um lote com $360m^2$ de área
- Em termos estatísticos, tanto faz utilizar uma variável explicativa x ou uma transformação linear ($x_{center} = (x-a)/b$) dela
    - De fato, é muito comum entre os estatísticos fazer, para cada X_i:
        - $$X_{ic} = \frac{X_i - \overline{X}_i}{SD(X_i)}$$
        - Isto é conhecido como padronização de variáveis!
    - R^2 será o mesmo, os $\hat{\beta}_i$ serão os mesmos, apenas $\hat{\beta}_0$ se modifica

Cobb-Douglas (7)



- $a = \exp(8,93) \approx 7.560,00;\ b \approx -0,25$
- $PU = 7.560 \cdot \left(\frac{Area}{360}\right)^{-0,25} = 7.560 \cdot \left(\frac{360}{Area}\right)^{+0,25}$

Derivação do Fator Área

- $$PU = 7.560 \cdot \left(\frac{360}{Area}\right)^{+0{,}25} \qquad (3)$$
- Com o modelo da Equação 3 é fácil derivar:
  - O fator área (F_S):
    - $$F_S = \left(\frac{360}{Area}\right)^{0{,}25}$$
    - É fácil ver que:
      - Quando $Area = 360m^2, F_S = 1$
      - Quando $Area > 360$, então $F_S < 1,0$
      - Quando $Area < 360$, então $F_S > 1,0$
  - $\overline{PU}_{Hom} = \exp(\hat{\beta}_0)$ (se a variável explicativa estiver centralizada!)

Aplicação do Fator Área

- Para homogeneização:
  - Imóvel da amostra com $440m^2$ e $PU = \text{ R\$ } 7.000/m^2$:
    - $F_S = (360/440)^{.25} \approx 0,95$
    - $PU_{Hom} = 7.000/0,95 \approx \text{ R\$ } 7.368,00/m^2$
- Para a avaliação:
  - Imóvel avaliando com $250m^2$, em amostra com $\overline{PU}_{Hom} = 7.560$:
    - $F_S = (360/250)^{0,25} = 1,0955$
    - $PU_{aval} = 7.560 \cdot 1,0955 \approx 8.280,00$

Múltiplos Fatores Multiplicativos

Cobb-Douglas

- É muito bem difundida entre os economistas a seguinte função de produção de *Cobb-Douglas*:
  - $Q(L, K) = A \cdot L^{\beta} \cdot K^{\alpha}$
    - $A > 0$ é uma constante
    - L é o fator Trabalho
    - K é o fator capital
    - α e β são coeficientes que dependem da indústria
    - Q é a quantidade produzida, função de L e K
- Como fazem os economistas para estimar A, α e β:
  - $$\begin{aligned} \ln(Q) &= \ln(A \cdot L^{\beta} \cdot K^{\alpha}) \\ \ln(Q) &= \ln(A) + \ln(L^{\beta}) + \ln(K^{\alpha}) \\ \ln(Q) &= \ln(A) + \beta \ln(L) + \alpha \ln(K) \end{aligned}$$

Cobb-Douglas

- Da mesma forma, portanto, temos um modelo de regressão do tipo:

$$\widehat{\ln(PU)} = \hat{\beta}_0 + \hat{\beta}_1 \ln(L) + \hat{\beta}_2 \ln(K)$$

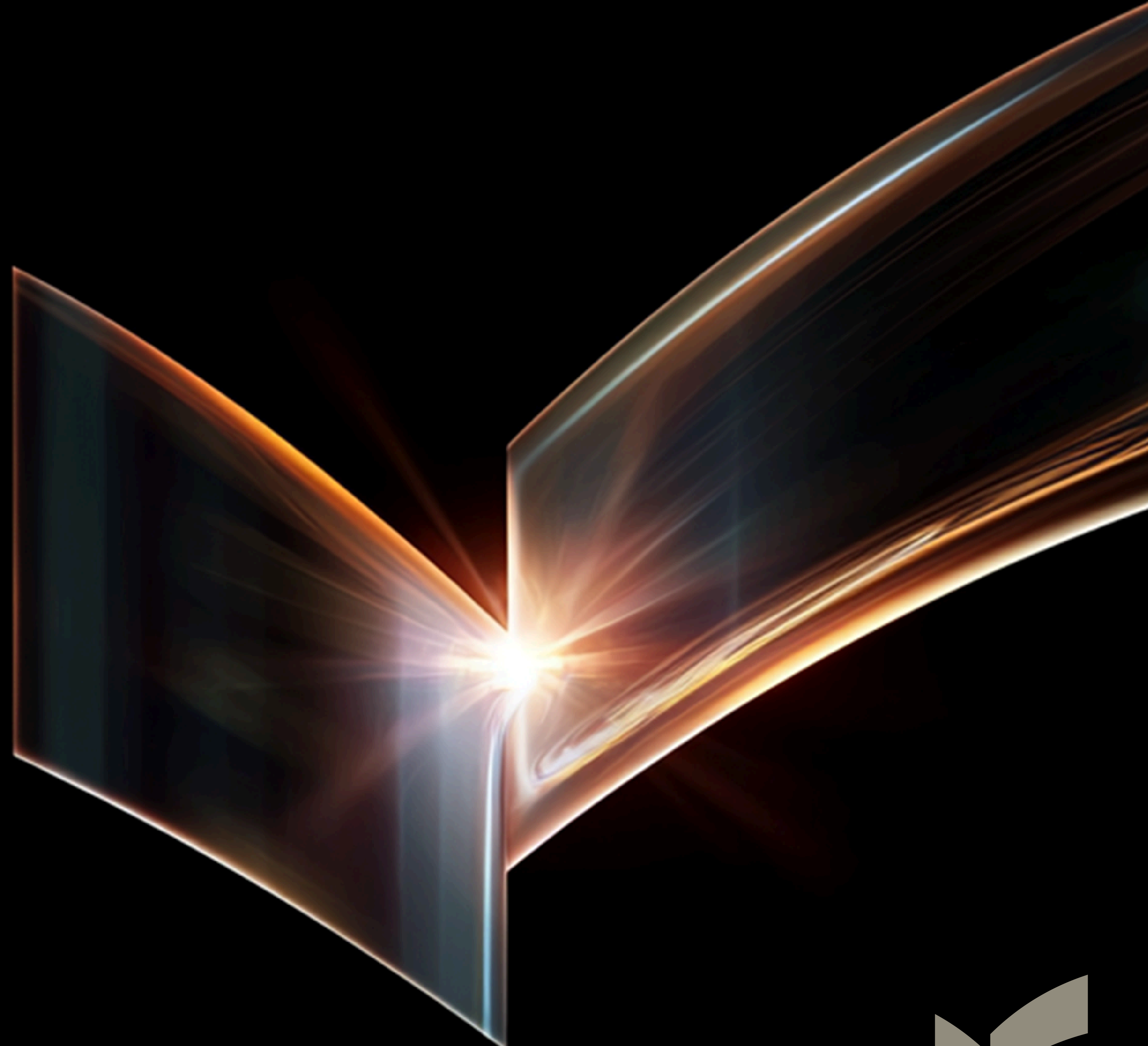
- Uma vez estimados $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$, temos:
  - $A = \exp(\hat{\beta}_0)$
  - $\beta = \hat{\beta}_1$
  - $\alpha = \hat{\beta}_2$

Derivação de Fatores em Modelos RLM

- Nos modelos multiplicativos de regressão linear múltipla (**RLM**), temos:
  - $\ln(PU) = \beta_0 + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) + \ldots + \beta_k \ln(X_k) + \varepsilon$
- Centralizando estes modelos, teremos:
  - $\ln(PU) = \beta_0^* + \beta_1 \ln\left(\frac{X_1}{c_1}\right) + \beta_2 \ln\left(\frac{X_2}{c_2}\right) + \ldots + \beta_k \ln\left(\frac{X_k}{c_k}\right) + \varepsilon$
- Estimando o modelo da equação acima, obtemos a eq. de Estimação:
  - $\widehat{PU} = \exp(\hat{\beta}_0^*) \cdot (X_1/c_1)^{\hat{\beta}_1} \cdot (X_2/c_2)^{\hat{\beta}_2} \cdot \ldots \cdot (X_k/c_k)^{\hat{\beta}_k} \cdot \exp(\varepsilon)$

- $\overline{PU}_{Hom} = \exp(\hat{\beta}_0^*)$
- $F_1 = (X_1/c_1)^{\hat{\beta}_1}$
- $F_2 = (X_2/c_2)^{\hat{\beta}_2}$
- $\ldots$
- $F_k = (X_k/c_k)^{\hat{\beta}_k}$

Exemplo 1



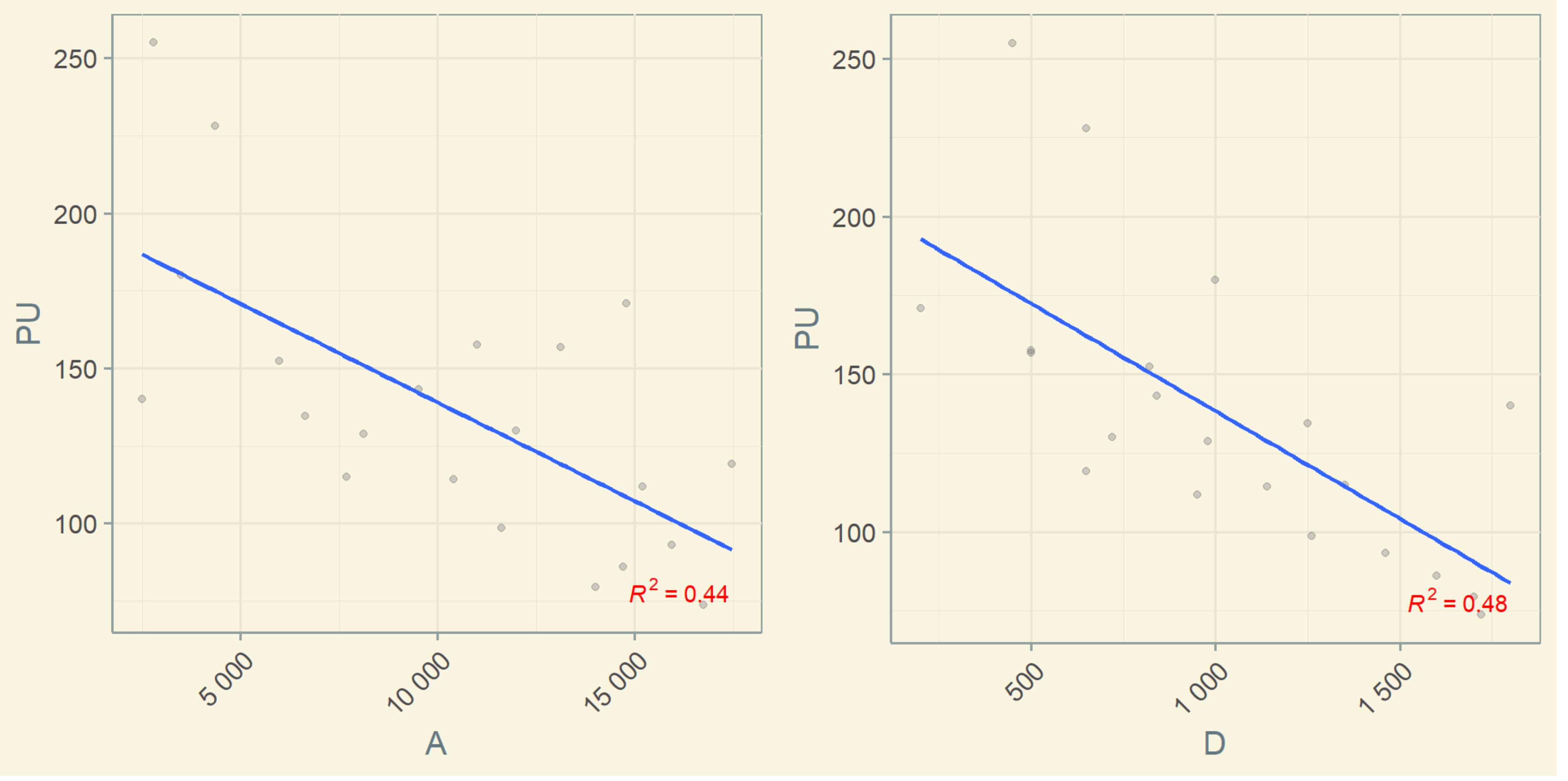
Dados

- Zeni (2024):

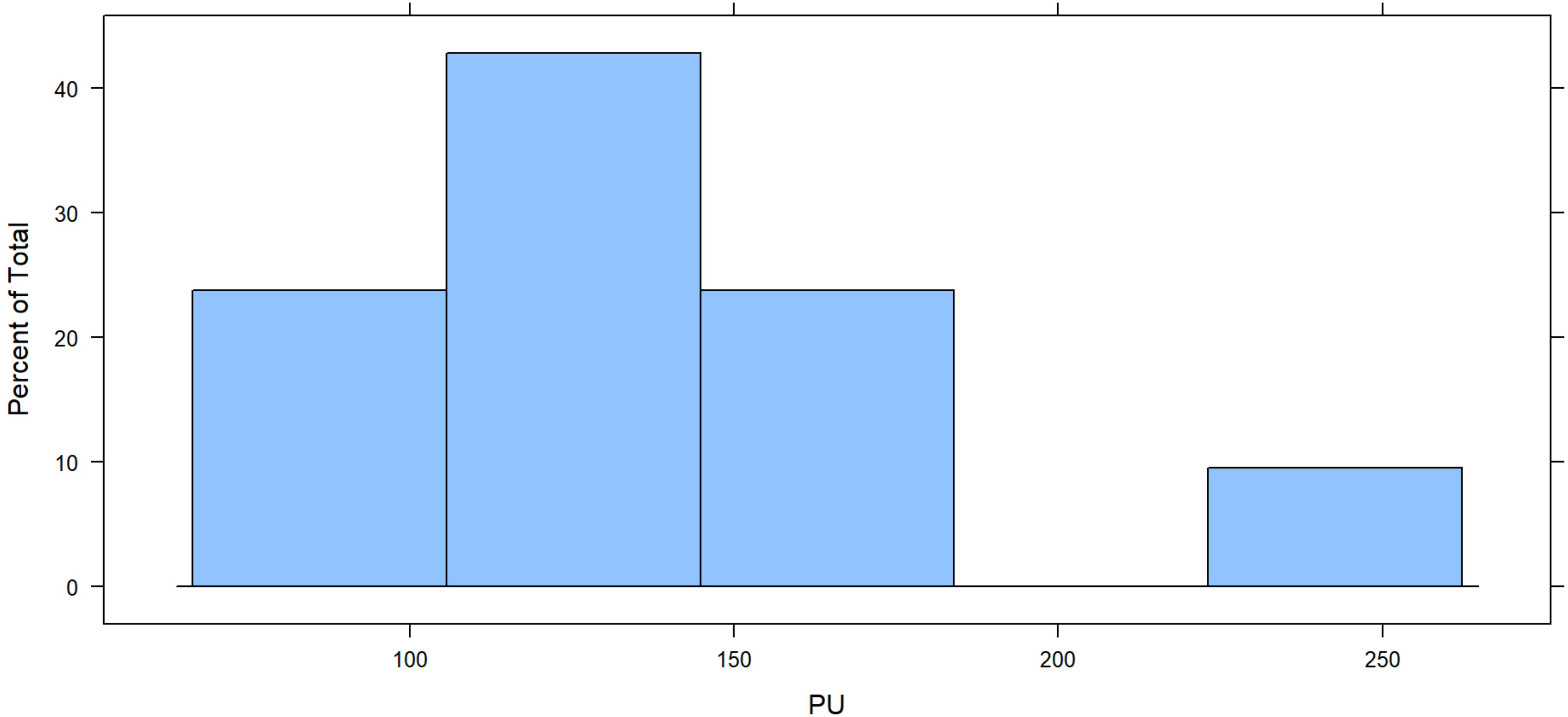
Id	A	D	PU
1	2.500	1.800	140,00
2	2.800	450	255,00
3	3.500	1.000	180,00
4	4.350	650	228,00
5	5.980	820	152,30
6	6.650	1.250	134,50
7	7.700	1.350	115,00
8	8.120	980	128,75
9	9.520	840	143,10
10	10.410	1.140	114,25
11	11.000	500	157,50

Id	A	D	PU
12	11.630	1.260	98,65
13	12.000	720	130,00
14	13.120	500	156,70
15	14.000	1.700	79,50
16	14.700	1.600	86,00
17	15.200	950	111,90
18	15.940	1.460	93,20
19	16.750	1.720	73,80
20	17.470	650	119,30
21	14.800	200	171,00

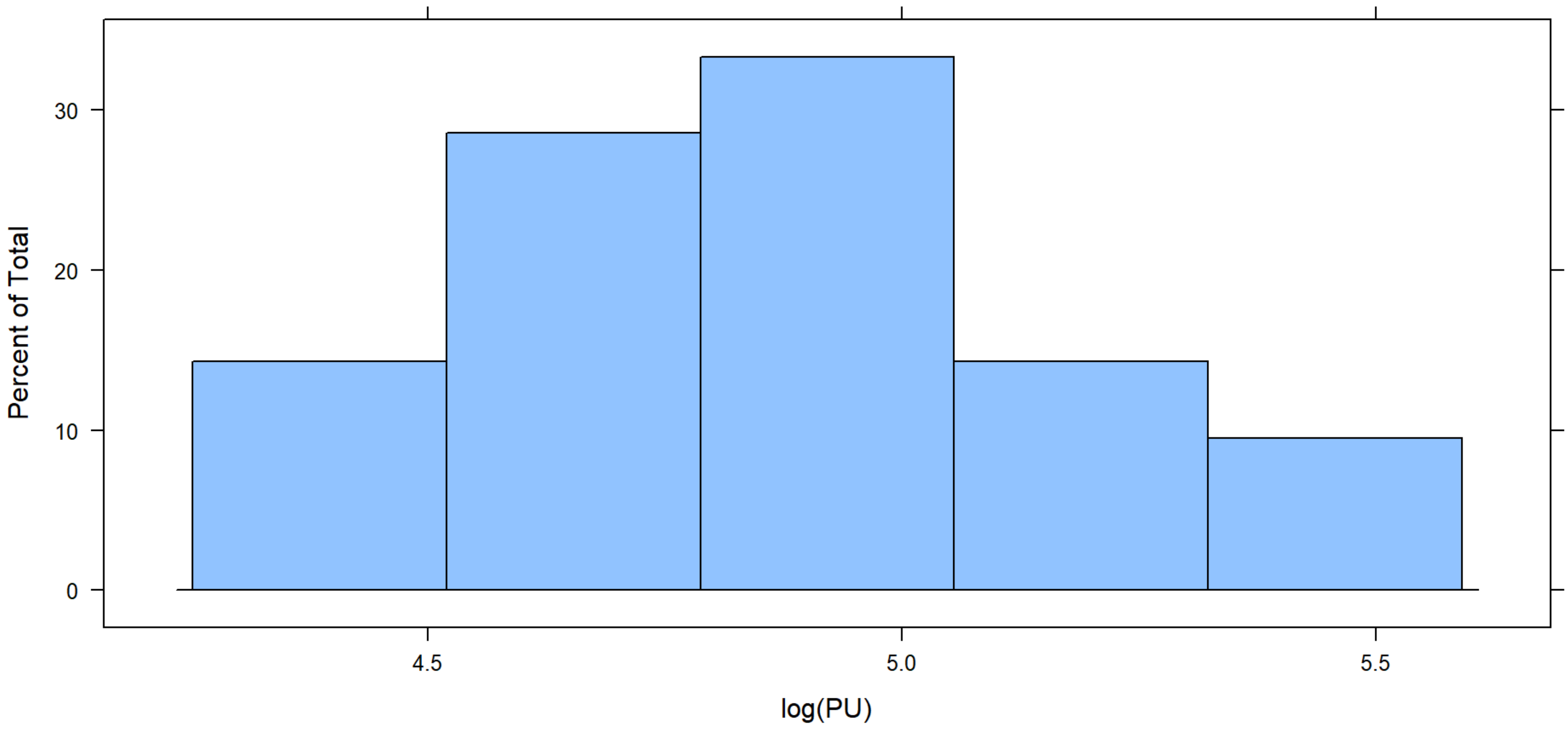
Análise Exploratória



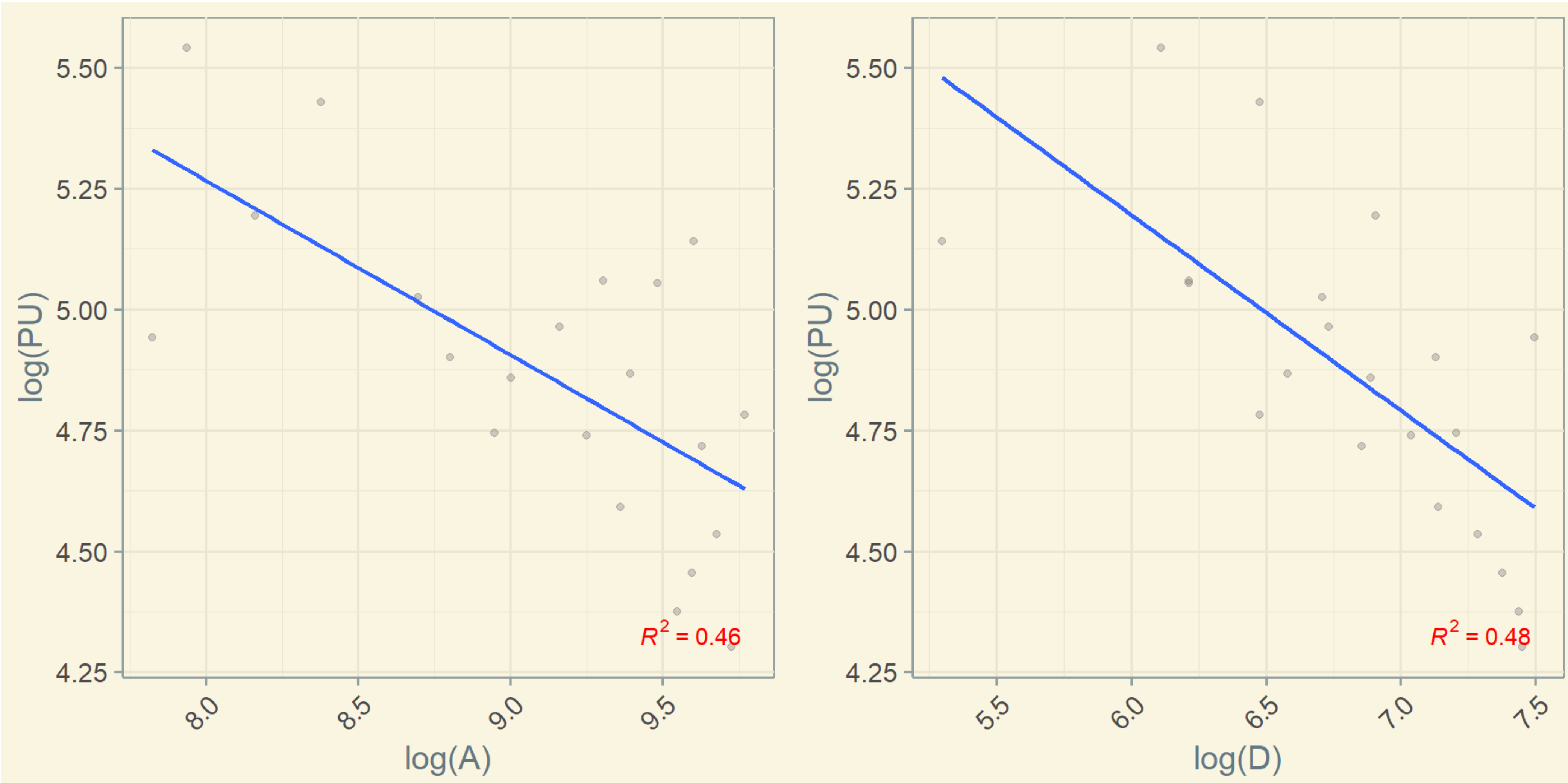
Análise Exploratória



Análise Exploratória



Análise Exploratória



Ajuste do Modelo

```
Call: lm(formula = log(PU) ~ log(A) + log(D), data = zeni_2024a)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.95401    0.34272   31.96  < 2e-16 ***
log(A)      -0.36349    0.02901  -12.53 2.50e-10 ***
log(D)      -0.40761    0.03164  -12.88 1.59e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.07774 on 18 degrees of freedom
Multiple R-squared: 0.9468
F-statistic: 160.2 on 2 and 18 DF,  p-value: 3.413e-12
   AIC    BIC
-42.92 -38.75
```

Aparte: Resíduos Parciais

- Em um modelo de regressão linear, temos:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$$

- Os resíduos totais, portanto, são:

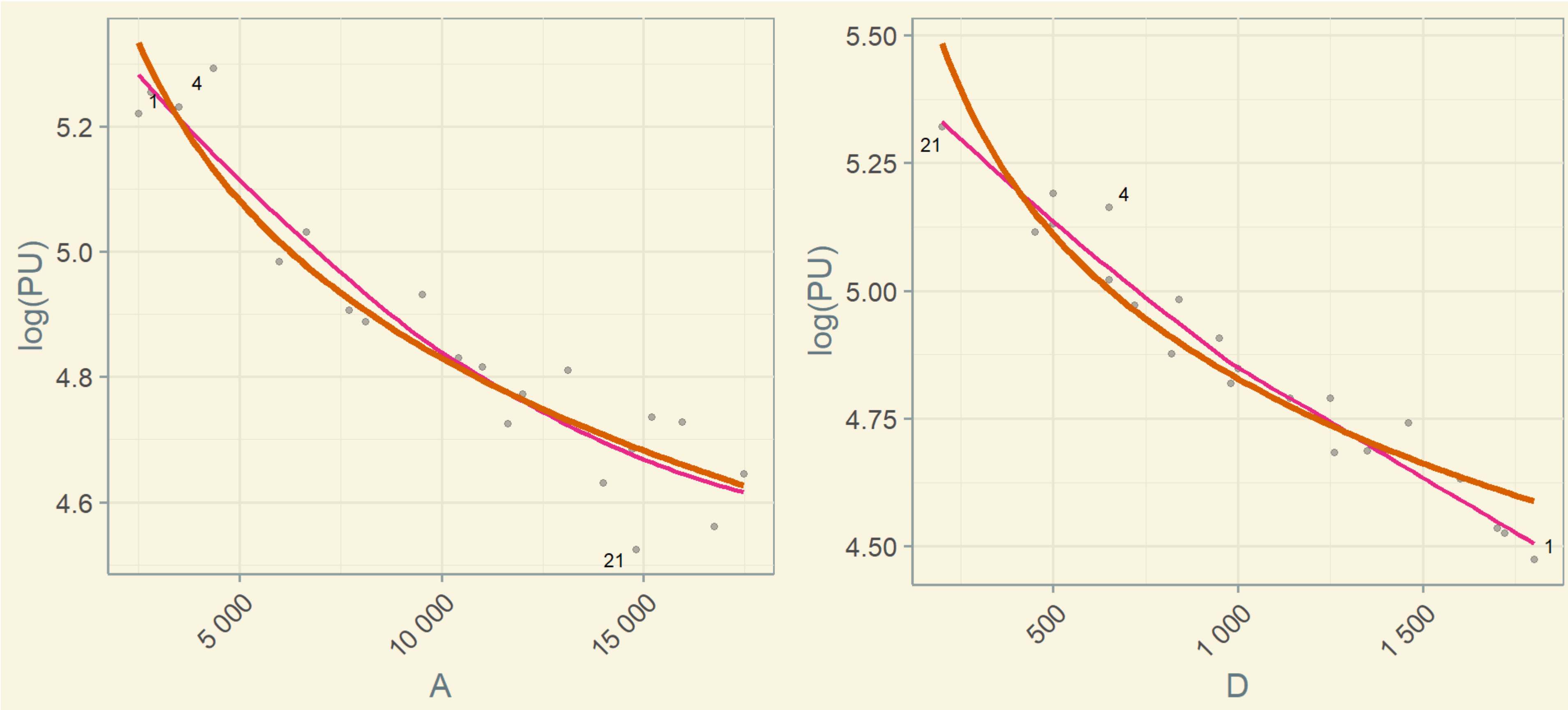
$$\hat{\varepsilon} = Y - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k)$$
$$\hat{\varepsilon} = Y - \hat{Y}$$

- Para obter resíduos parciais em relação a uma variável explicativa, retornamos o efeito dela à equação acima:
  - $\hat{\varepsilon}_1 = Y - \hat{Y} + \hat{\beta}_1 X_1$
  - $\hat{\varepsilon}_2 = Y - \hat{Y} + \hat{\beta}_2 X_2$
  - …
  - $\hat{\varepsilon}_k = Y - \hat{Y} + \hat{\beta}_k X_k$

Plotagem do Modelo

- Os resíduos parciais são úteis para serem plotados contra as variáveis explicativas:

Gráficos de Resíduos Parciais



log(PU) ~ log(A) + log(D)

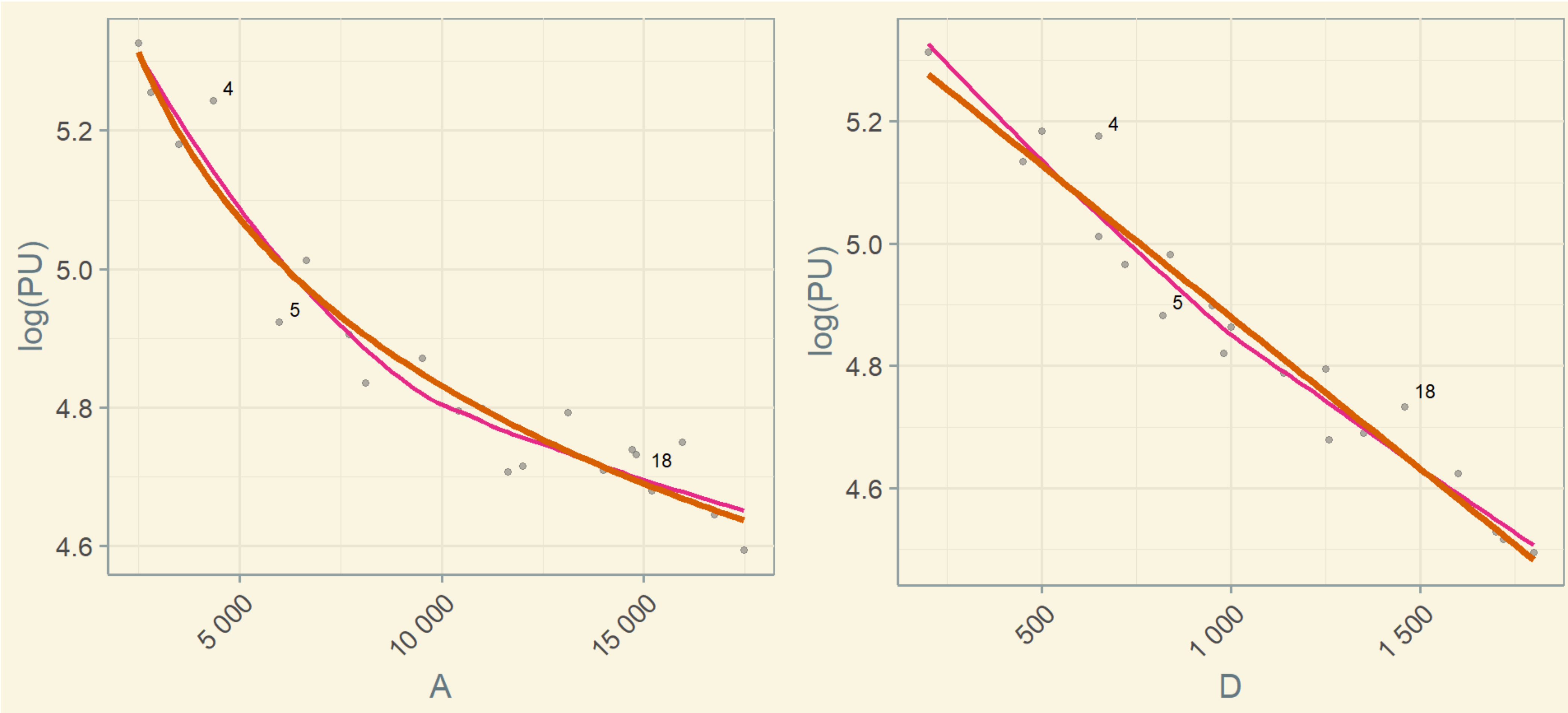
Novo Modelo

```
Call: lm(formula = log(PU) ~ log(A) + D, data = zeni_2024a)

Coefficients:
               Estimate  Std. Error t value Pr(>|t|)
(Intercept)  8.54331250  0.18625876   45.87  < 2e-16 ***
log(A)      -0.34773547  0.02029311  -17.14 1.36e-12 ***
D           -0.00049622  0.00002621  -18.93 2.48e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.05436 on 18 degrees of freedom
Multiple R-squared: 0.974
F-statistic:   337 on 2 and 18 DF,  p-value: 5.453e-15
   AIC    BIC
-57.95 -53.77
```

- Note-se o aumento do grau de ajuste do modelo!

Plotagem do Novo Modelo

Gráficos de Resíduos Parciais



log(PU) ~ log(A) + D

Centralização dos Dados

- $A = 11.000$ (mediana da variável Área)
- $D = 980$ (moda da variável Distância)

Modelo Centralizado

```
Call: lm(formula = log(PU) ~ log(A/11000) + I(D - 980), data = zeni_2024a)

Coefficients:
                Estimate  Std. Error t value Pr(>|t|)
(Intercept)   4.82110852  0.01259858  382.67  < 2e-16 ***
log(A/11000) -0.34773547  0.02029311  -17.14 1.36e-12 ***
I(D - 980)   -0.00049622  0.00002621  -18.93 2.48e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.05436 on 18 degrees of freedom
Multiple R-squared: 0.974
F-statistic:   337 on 2 and 18 DF,  p-value: 5.453e-15
   AIC    BIC
-57.95 -53.77
```

- $\overline{PU}_{Hom} = \exp(4,82) \approx \text{ R\$ } 124,0/m^2$

Derivação de Fatores

- Fator Área (F_S):
- $F_S = (11.000/A)^{0,35}$
- Fator Distância (F_D):
- $F_D = \exp(-5.10^{-3}(D - 980)) = 0,9995^{(D-980)}$

- Qual o valor de um lote com 10.000m^2 situado a uma distância de 100m?

- Previsão a partir dos fatores:
- $F_S = (11.000/A_i)^{0,35}$
- $F_S = (11.000/10.000)^{0,35}$
- $F_D = 0,9995^{(D_i-980)}$
- $F_D = 0,9995^{(100-980)}$
- $\widehat{PU} = \overline{PU}_{Hom} \cdot F_S \cdot F_D$
- $\widehat{PU} = 124,0 \cdot 1,034 \cdot 1,55$
- $\widehat{PU} \approx \text{R\$ } 200,0/m^2$

- Previsão a partir do modelo:
- $\widehat{\ln(PU)} = 4,82 - 0,35\ln(,91) - 5.10^{-3}.880$
- $\widehat{\ln(PU)} = 5{,}29$
- $\widehat{PU} \approx \exp(5,30)$
- $\widehat{PU} \approx \text{R\$ } 200,0/m^2$
- OK!

Outros tipos de variáveis explicativas

Outras variáveis

- Variáveis dicotômicas
  - As variáveis dicotômicas são variáveis do tipo 0/1
    - Usualmente: não (0)/ sim (1)
    - Exemplo:
      - Prédio com elevador? não/sim (0/1)
- As variáveis dicotômicas não sofrem transformações
- Um modelo que inclui uma variável dicotômica (X_2) será do tipo:
  - $$\ln(PU) = \beta_0 + \beta_1 \ln(X_1) + \beta_2 X_2 + \varepsilon$$

Outras variáveis (2)

- Um modelo que inclui uma variável dicotômica (X_2) será do tipo:
  - $$\ln(PU) = \beta_0 + \beta_1 \ln(X_1) + \beta_2 X_2 + \varepsilon$$
  - Portanto, a sua equação de estimação será:
    - $$\hat{PU} = \exp(\hat{\beta}_0) \cdot X_1^{\hat{\beta}_1} \cdot \exp(\hat{\beta}_2 \cdot X_2)$$
    - Como $\exp(k.\,X) = \exp(k)^X \rightarrow \exp(\hat{\beta}_2 \cdot X_2) = \exp(\hat{\beta}_2)^{X_2}$, então:
    - $$PU = \exp(\hat{\beta}_0) \cdot X_1^{\hat{\beta}_1} \cdot \exp(\hat{\beta}_2)^{X_2}$$
      - Ou seja, se:
        - $X_2 = 0 \rightarrow \exp(\hat{\beta}_2)^{X_2} = \exp(\hat{\beta}_2)^0 = 1$
        - $X_2 = 1 \rightarrow \exp(\hat{\beta}_2)^{X_2} = \exp(\hat{\beta}_2)^1 = \exp(\hat{\beta}_2)$

Outras variáveis (3)

- No caso de variáveis numéricas contínuas, sem transformação, tais como:
  - Quartos (Q), com valores possíveis igual a $\{1, 2, 3, 4, \ldots\}$;
  - Vagas de Garagens (G), com valores possíveis igual a $\{0, 1, 2, 3, \ldots\}$;
  - Padrão Construtivo (PC) com códigos alocados ou dicotomia em grupo;
  - Vale o mesmo raciocínio das variáveis dicotômicas!
- Exemplo - $X_1 = \text{Area}, X_2 = \text{Elevador}, X_3 = G$:
  - $$PU = \exp(\hat{\beta}_0) \cdot \exp(\hat{\beta}_2)^{\text{Elevador}} \cdot \exp(\hat{\beta}_3)^G$$
  - Se $G = 0 \rightarrow \exp(\hat{\beta}_3)^0 = 1$
  - Se $G = 1 \rightarrow \exp(\hat{\beta}_3)^1 = \exp(\hat{\beta}_3)$
  - Se $G = 2 \rightarrow \exp(\hat{\beta}_3)^2$
  - Se $G = n \rightarrow \exp(\hat{\beta}_3)^n$

Outras variáveis (4)

- E se ao ajustar o modelo **RLM** for utilizada a transformação $\sqrt{G}$ (ou G^2, ou qualquer outra transformação, excetuando-se $\ln(G)$):
  - $$PU = \exp(\hat{\beta}_0) + \exp(\hat{\beta}_2)^{\text{Elevador}} + \exp(\hat{\beta}_3)^{\sqrt{G}}$$
  - Se $G = 0 \to \exp(\hat{\beta}_3)^{\sqrt{0}} = 1$
  - Se $G = 1 \to \exp(\hat{\beta}_3)^{\sqrt{1}} = \exp(\hat{\beta}_3)$
  - Se $G = 2 \to \exp(\hat{\beta}_3)^{\sqrt{2}} = \exp(\hat{\beta}_3)^{1,414}$
  - Se $G = n \to \exp(\hat{\beta}_3)^{\sqrt{n}} = \exp(\hat{\beta}_3)^{1,732}$

Outras variáveis (5)

- Para a dicotomia em grupo, vale tudo que se aplica à dicotomia simples:
  - Não cabem transformações das variáveis dicotômicas em grupo!
  - Os fatores serão obtidos através da exponenciação dos coeficientes
  - A variável passa a ser expoente de $\exp(\hat{\beta}_i)$

Outras variáveis (6)

- Assim como é comum utilizar variáveis dicotômicas para testar a alteração dos níveis de preços entre um período e outro, ou entre um bairro e outro, etc.
- É possível utilizar a *dicotomia em grupo* para testar a alteração dos níveis de preços entre 3 ou mais períodos, ou 3 ou mais bairros, etc.
- Na dicotomia em grupo são criadas $g-1$ variáveis para simular os g grupos existentes.
  - Quanto há apenas 2 grupos, cria-se apenas uma variável e a dicotomia em grupo degenera para a dicotomia simples
  - Quando há 3 grupos, deve-se criar 2 variáveis dicotômicas
  - Quando há 4 grupos, deve-se criar 3 variáveis dicotômicas
  - Etc.

Outras variáveis (7)

- Na dicotomia em grupo, portanto, teremos $g - 1$ coeficientes que medem a diferença de preços entre o grupo de referência e o grupo em análise.
- Por exemplo:
  - A variável *PC* representa o Padrão Construtivo, com 3 níveis:
    - *B* para Padrão Baixo;
    - *M* para Padrão Médio;
    - *A* para Padrão Alto.
  - O nível de referência pode ser qualquer um dos três níveis. Por exemplo, se escolhemos como referência o nível *B*, teremos duas variáveis explicativas:
    - *PCM*, que irá representar os imóveis de Padrão Médio e;
    - *PCA*, que irá representar os imóveis de Padrão Alto;
- Na matriz do modelo, estas duas variáveis serão dicotômicas (0/1).

Outras variáveis (8)

- Ex. (Zilli 2020):

```
   (Intercept) PCMédio PCAlto NG DABM  AP
1            1       0      1  1  367  97
2            1       1      0  2  807  73
3            1       1      0  2  803 120
4            1       0      1  1  136 122
5            1       1      0  1  116 110
6            1       0      0  1  143  84
7            1       1      0  1  319  95
8            1       0      1  1  892  54
9            1       1      0  2  585 138
10           1       1      0  1  660  60
11           1       0      1  2  282  96
12           1       1      0  2  475 120
13           1       0      1  1  338 120
14           1       0      1  3  528 167
15           1       1      0  1  358  67
16           1       1      0  2  674 109
17           1       0      1  1   91 110
18           1       0      1  2  330 159
```

```
Call: lm(formula = PU ~ PC + NG + DABM + AP, data = Zilli2020)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5661.9162   404.4109  14.000  < 2e-16 ***
PCMédio     1575.5161   285.8764   5.511 9.97e-08 ***
PCAlto      3359.2442   305.4220  10.999  < 2e-16 ***
NG          1709.5457   191.6150   8.922  < 2e-16 ***
DABM          -1.1247     0.2157  -5.214 4.26e-07 ***
AP           -12.0010     3.0807  -3.896  0.00013 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 1538 on 219 degrees of freedom
Multiple R-squared: 0.6741
F-statistic: 90.61 on 5 and 219 DF,  p-value: < 2.2e-16
    AIC     BIC
3948.68 3972.60
```

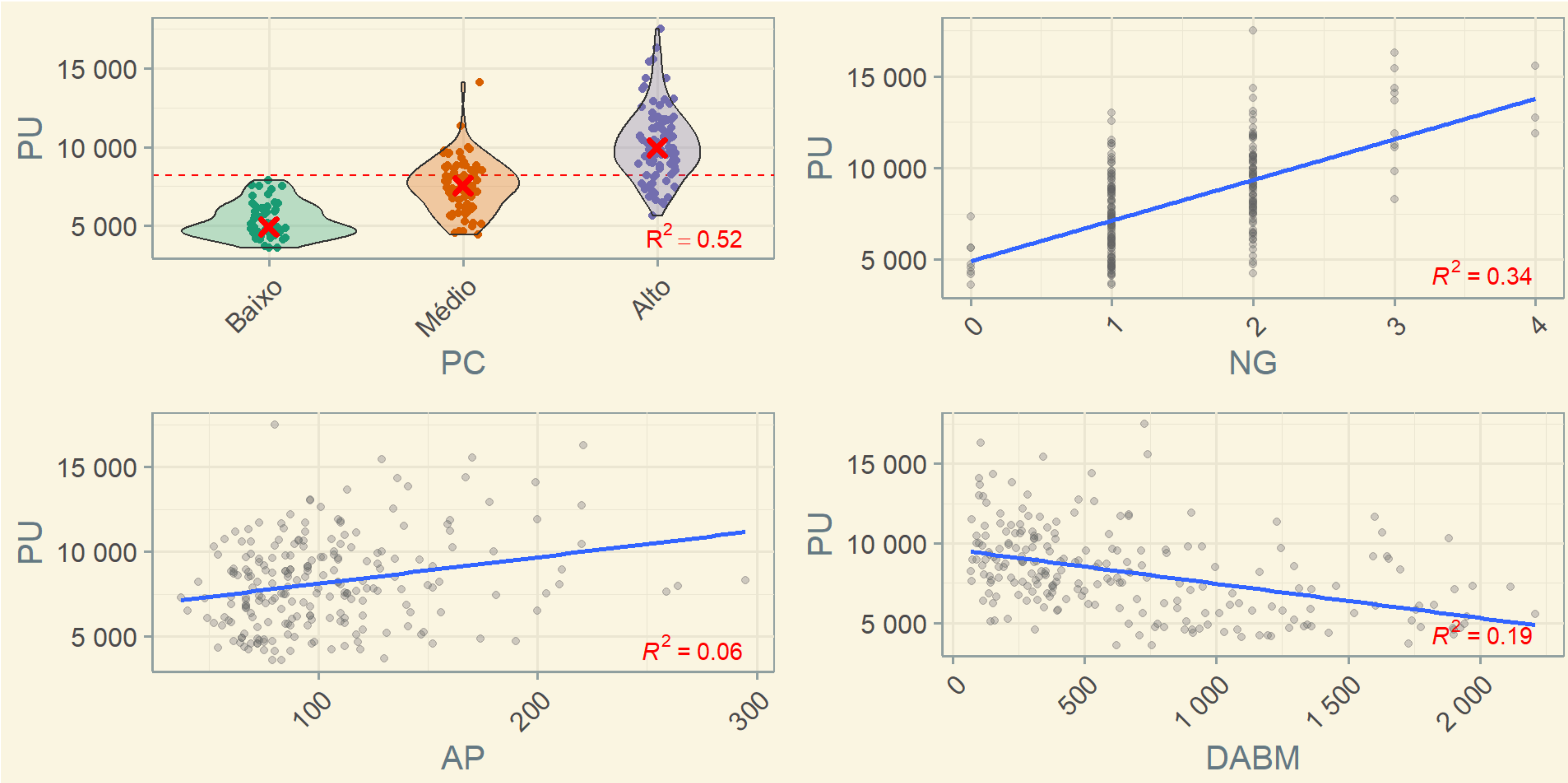
- Coeficientes das variáveis dicotômicas representam a **diferença** entre o *PU* do *PC* de referência (Baixo) em relação aos *PCM* e *PCA*!

Exemplo 2

Dados

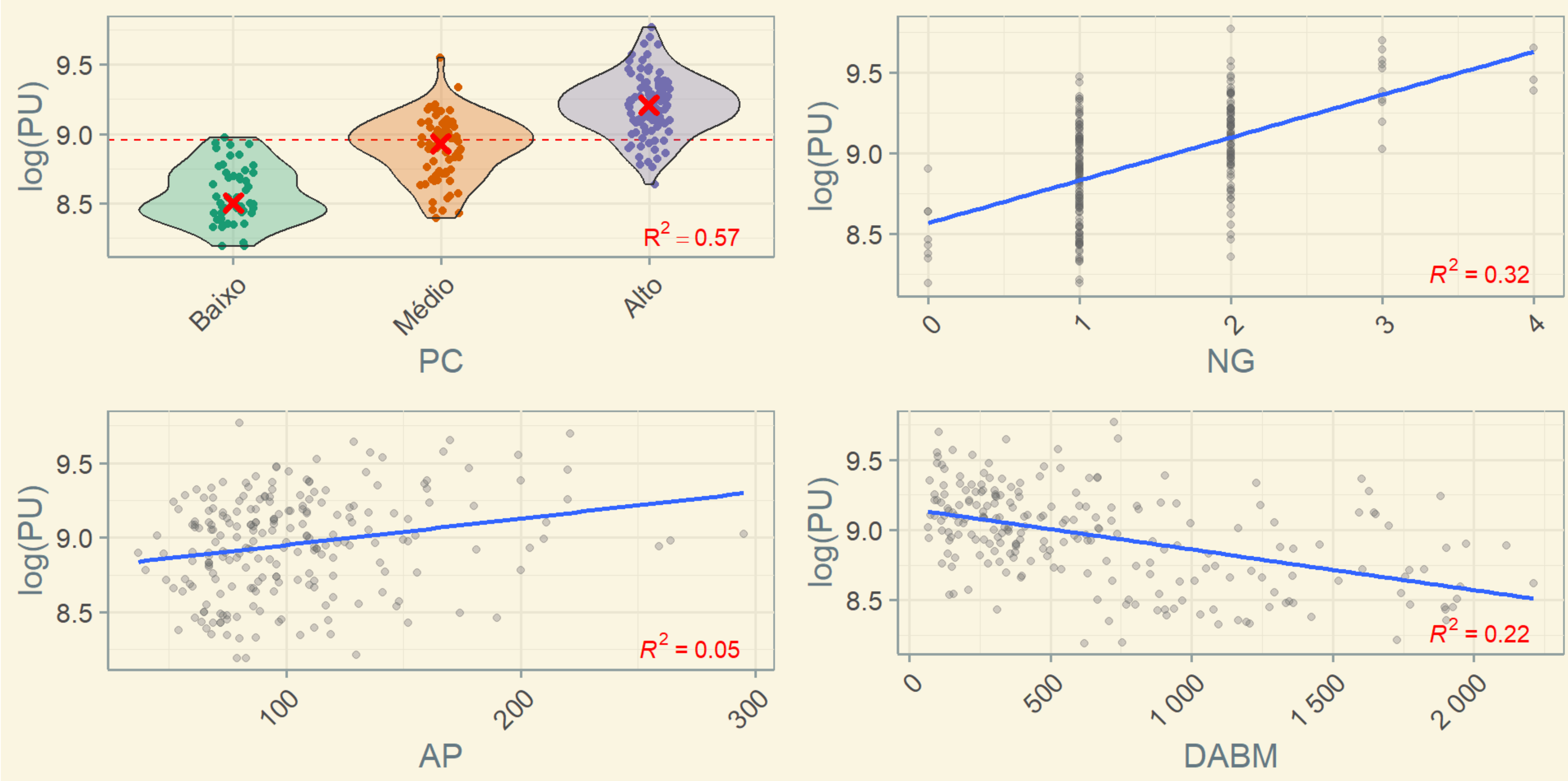
id	PT	PU	AP	PC	NG	NB	ND	NS	MO	PSN	CH	DABM	DSBM
1	920.045	9.485	97	Alto	1	3	3	1	MO	Sim	Sim	367	876
2	699.997	9.589	73	Médio	2	3	2	2	N	Sim	Sim	807	1.266
3	750.000	6.250	120	Médio	2	4	2	2	N	Nao	Nao	803	1.324
4	1.149.972	9.426	122	Alto	1	3	3	1	SM	Nao	Nao	136	1.064
5	700.040	6.364	110	Médio	1	2	3	1	MO	Nao	Nao	116	936
6	630.000	7.500	84	Baixo	1	2	2	0	N	Nao	Nao	143	1.668
7	839.990	8.842	95	Médio	1	2	3	1	MO	Sim	Sim	319	2.615
8	530.010	9.815	54	Alto	1	1	3	0	MO	Nao	Nao	892	1.971
9	1.075.020	7.790	138	Médio	2	4	3	1	MO	Nao	Nao	585	1.291
10	529.980	8.833	60	Médio	1	2	2	1	MO	Nao	Nao	660	1.164

Análise Exploratória



- Reparem na relação entre *PU* e *AP*!

Análise Exploratória

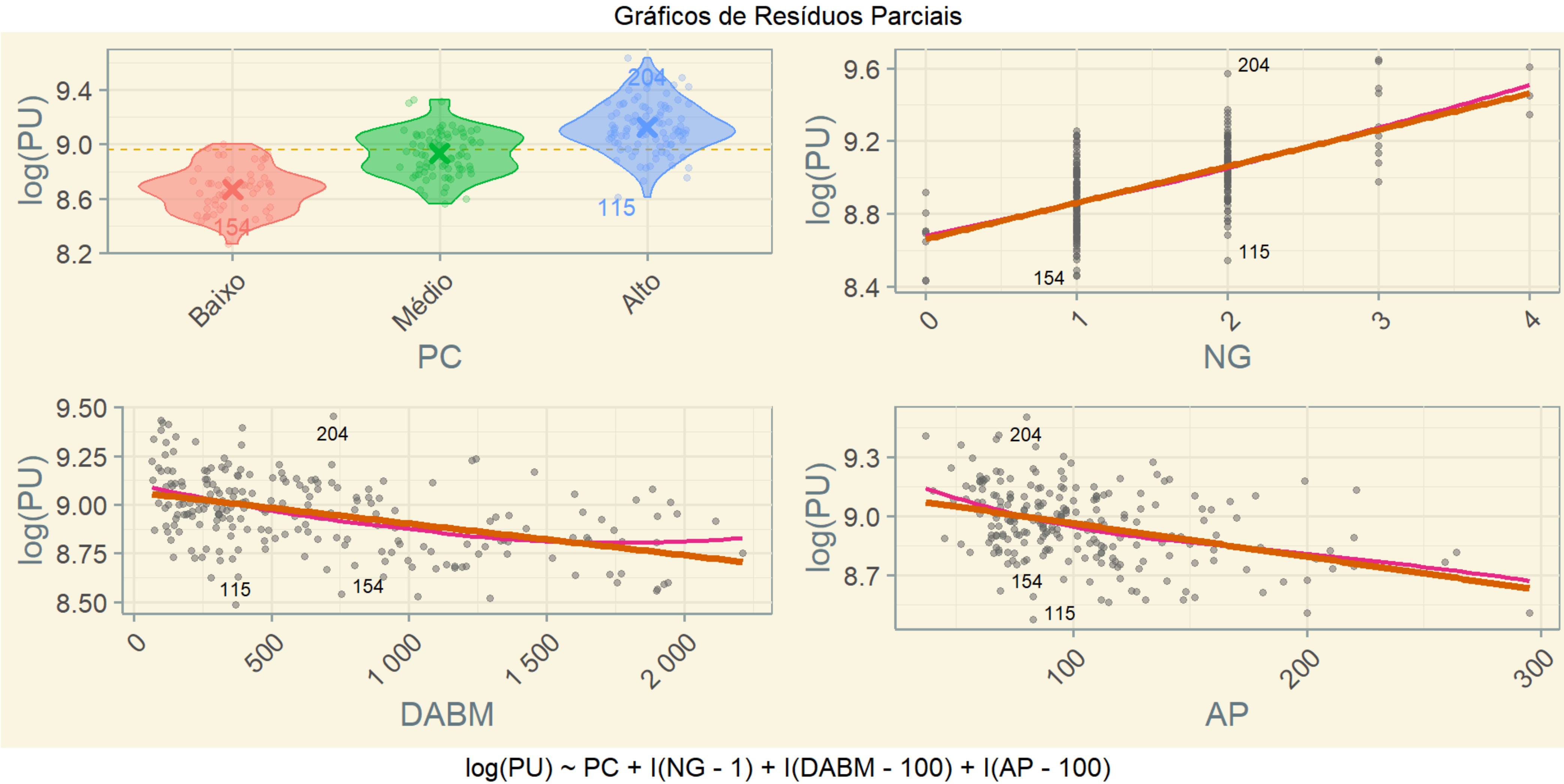


- Reparem na relação entre $\ln(PU)$ e *AP*!

Ajuste do Modelo RLM

```
Call: lm(formula = log(PU) ~ PC + I(NG - 1) + I(DABM - 100) + I(AP - 100), data
         = Zilli2020)

Coefficients:
                 Estimate  Std. Error t value Pr(>|t|)
(Intercept)    8.66956323  0.03183323 272.343  < 2e-16 ***
PCMédio        0.26553039  0.03272877   8.113 3.51e-14 ***
PCAlto         0.46075089  0.03496646  13.177  < 2e-16 ***
I(NG - 1)      0.20003869  0.02193718   9.119  < 2e-16 ***
I(DABM - 100) -0.00016208  0.00002469  -6.564 3.75e-10 ***
I(AP - 100)   -0.00169759  0.00035270  -4.813 2.77e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.1761 on 219 degrees of freedom
Multiple R-squared: 0.7224
F-statistic:   114 on 5 and 219 DF,  p-value: < 2.2e-16
    AIC     BIC
-135.09 -111.18
```

Plotagem do Modelo



- Reparem em como a relação entre *AP* e $\ln(PU)$ se modificou!

Derivação do Fator *PC*

```
Coefficients:
                 Estimate  Std. Error t value Pr(>|t|)
(Intercept)    8.66956323  0.03183323 272.343  < 2e-16 ***
PCMédio        0.26553039  0.03272877   8.113 3.51e-14 ***
PCAlto         0.46075089  0.03496646  13.177  < 2e-16 ***
I(NG - 1)      0.20003869  0.02193718   9.119  < 2e-16 ***
I(DABM - 100) -0.00016208  0.00002469  -6.564 3.75e-10 ***
I(AP - 100)   -0.00169759  0.00035270  -4.813 2.77e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.1761 on 219 degrees of freedom
Multiple R-squared: 0.7224
F-statistic:   114 on 5 and 219 DF,  p-value: < 2.2e-16
    AIC     BIC
-135.09 -111.18
```

- $\exp(\hat{\beta}_{PCM}) = \exp(0,2655) \approx 1,30$
- $\exp(\hat{\beta}_{PCA}) = \exp(0,46075) \approx 1,59$
- $F_{PC} = \begin{cases} 1,00 & \text{se } PC = \text{Baixo} \\ 1,30 & \text{se } PC = \text{Médio} \\ 1,59 & \text{se } PC = \text{Alto} \end{cases}$

Paradoxo de Simpson

Paradoxo de Simpson

- Como vimos no exemplo anterior:
  - A relação entre PU e AP era uma relação fraca e positiva!
  - Porém, com o ajuste da *RLM*, essa relação tornou-se negativa e relevante!

- Correlações Parciais:

	Zero-order	**Partial**	**Part**
PCMédio	-0,12	0,48	0,29
PCAlto	0,66	0,67	0,47
I(NG - 1)	0,56	0,52	0,32
I(DABM - 100)	-0,47	-0,41	-0,23
I(AP - 100)	0,23	-0,31	-0,17

- Na tabela acima são vistas a correlação de ordem zero, a parcial e a semi-parcial (coluna *Part*).

Correlações Parciais

Correlations

Variable	Zero Order	Partial	Part
PCMédio	-0.119	0.481	0.289
PCAlto	0.658	0.665	0.469
I(NG - 1)	0.564	0.525	0.325
I(DABM - 100)	-0.467	-0.405	-0.234
I(AP - 100)	0.228	-0.309	-0.171

- O valor da correlação semi-parcial elevado ao quadrado é também conhecido como **coeficiente de determinação parcial**!
- Para a variável AP: $sr^2_{AP} = (-0,171)^2 \approx 0,03.$
- Para a variável DABM: $sr^2_{DABM} = (-0,234)^2 \approx 0,05.$
- Para a variável NG: $sr^2_{NG} = (-0,325)^2 \approx 0,10.$
- O coeficiente de determinação parcial de uma variável representa o percentual de explicação que ela **adiciona** ao modelo!

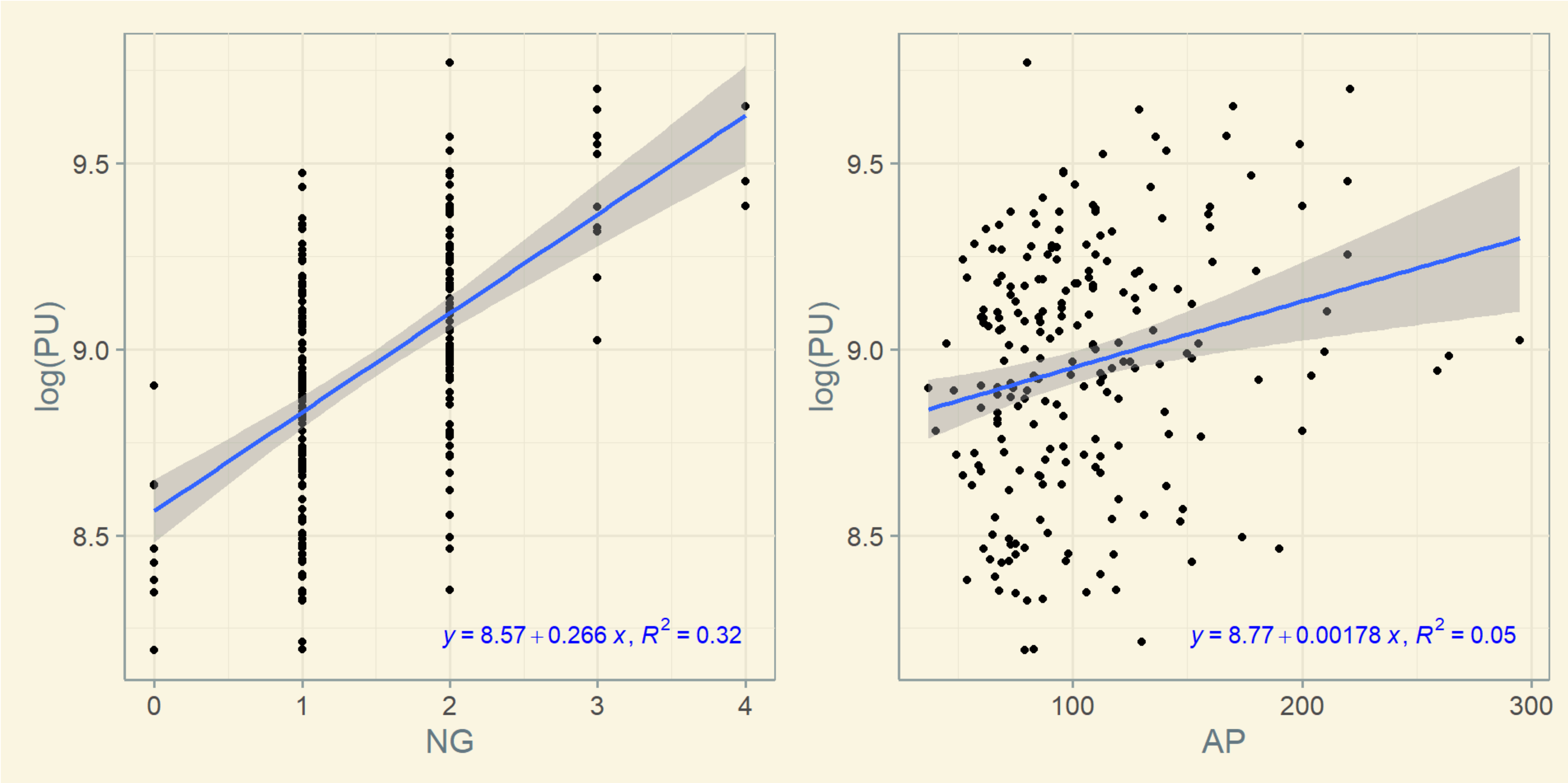
Correlações Parciais

```
Coefficients:
                 Estimate  Std. Error t value Pr(>|t|)    
(Intercept)    8.66495481  0.03338334 259.559  < 2e-16 ***
PCMédio        0.26618718  0.03433772   7.752 3.32e-13 ***
PCAlto         0.47655274  0.03652368  13.048  < 2e-16 ***
I(NG - 1)      0.14210373  0.01924175   7.385 3.11e-12 ***
I(DABM - 100) -0.00012863  0.00002486  -5.174 5.16e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.1848 on 220 degrees of freedom
Multiple R-squared: 0.693
F-statistic: 124.1 on 4 and 220 DF,  p-value: < 2.2e-16 
    AIC     BIC 
-114.47  -93.97 
```

```
Coefficients:
                 Estimate  Std. Error t value Pr(>|t|)    
(Intercept)    8.66956323  0.03183323 272.343  < 2e-16 ***
PCMédio        0.26553039  0.03272877   8.113 3.51e-14 ***
PCAlto         0.46075089  0.03496646  13.177  < 2e-16 ***
I(NG - 1)      0.20003869  0.02193718   9.119  < 2e-16 ***
I(DABM - 100) -0.00016208  0.00002469  -6.564 3.75e-10 ***
I(AP - 100)   -0.00169759  0.00035270  -4.813 2.77e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.1761 on 219 degrees of freedom
Multiple R-squared: 0.7224
F-statistic:   114 on 5 and 219 DF,  p-value: < 2.2e-16 
    AIC     BIC 
-135.09 -111.18 
```

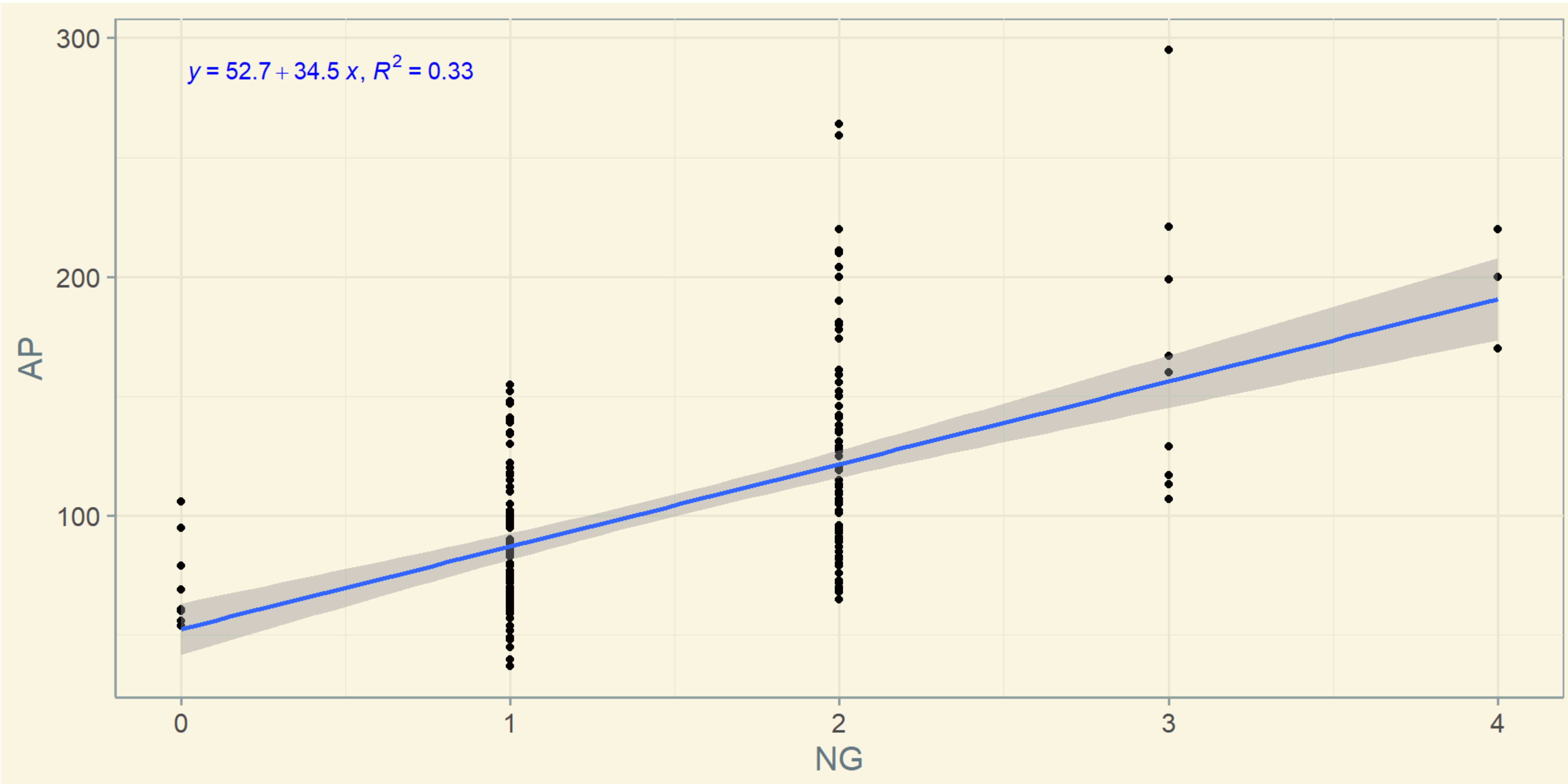
- Acréscimo de R^2 devido à inclusão de $AP = 0,7224 - 0,6934 \approx 0,03$

Por que muda a relação entre as variáveis?



Por que muda a relação entre as variáveis?

- Vejam a relação entre as variáveis explicativas AP e NG:



Por que muda a relação entre as variáveis?

- $\ln(PU) = \beta_{0NG} + \beta_{NG} \cdot \mathrm{NG} + \varepsilon_{NG}$
- $\ln(PU) = \beta_{0AP} + \beta_{AP} \cdot \mathrm{AP} + \varepsilon_{AP}$
- $\ln(PU) = \beta_{0C} + \beta_{1C} \cdot \mathrm{NG} + \beta_{2C} \cdot \mathrm{AP} + \varepsilon_{C}$
- $NG = \beta_{0X} + \beta_{1X} \cdot \mathrm{AP} + \varepsilon_{X}$
- Então:
  - $$\ln(PU) = \hat{\beta}_{0C} + \hat{\beta}_{1C}(\hat{\beta}_{0X} + \hat{\beta}_{1X})\mathrm{AP} + \hat{\beta}_{2C}\mathrm{AP}$$
    $$\ln(PU) = (\hat{\beta}_{0C} + \hat{\beta}_{1C}\hat{\beta}_{0X}) + (\hat{\beta}_{1C}\hat{\beta}_{1X} + \hat{\beta}_{2C})\mathrm{AP}$$
- $\hat{\beta}_{0AP} = \hat{\beta}_{0C} + \hat{\beta}_{1C}\hat{\beta}_{0X}$
- $\hat{\beta}_{AP} = \hat{\beta}_{1C}\hat{\beta}_{1X} + \hat{\beta}_{2C}$

Por que muda a relação entre as variáveis?

- $\ln(PU) = 8,6225 + 0,303 \cdot \mathrm{NG} - 0,00108\mathrm{AP}$
- $NG = 0,497 + 0,0094 \cdot \mathrm{AP}$
- $\beta_{0AP} = 8,6225 + 0,303 \cdot 0,497 \approx 8,77$
- $\beta_{AP} = 0,303 \cdot 0,0094 + (-0,00108) \approx +0,00177$

```
Call: lm(formula = log(PU) ~ AP, data = Zilli2020)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.7729213  0.0568735 154.253  < 2e-16 ***
AP          0.0017783  0.0005083   3.499 0.000564 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.3225 on 223 degrees of freedom
Multiple R-squared: 0.05204
F-statistic: 12.24 on 1 and 223 DF,  p-value: 0.0005641
   AIC    BIC
133.20 143.45
```

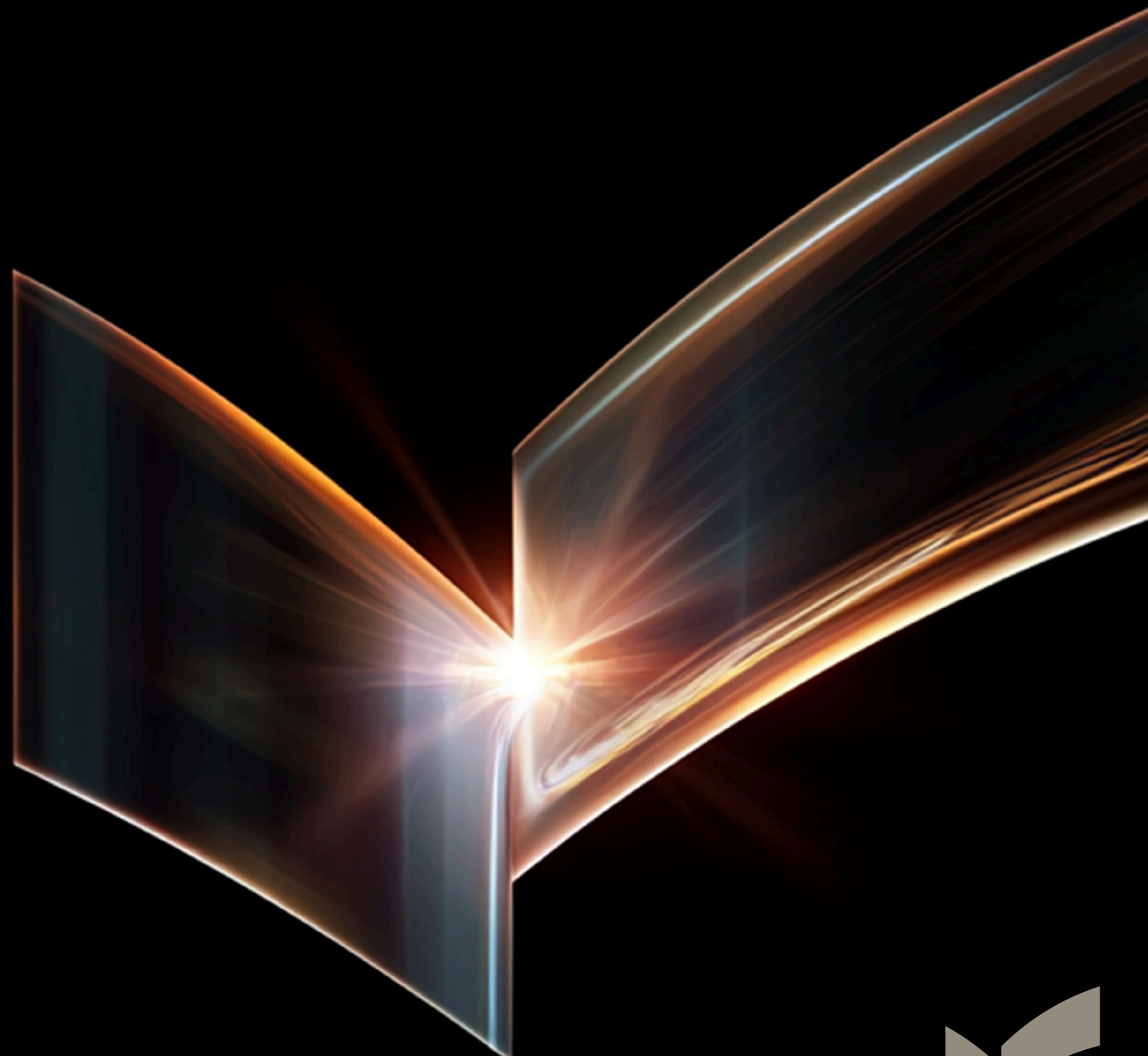
Por que muda a relação entre as variáveis?

- $\ln(PU) = 8,6225 + 0,303 \cdot \mathrm{NG} - 0,00108\mathrm{AP}$
- $AP = 52,7 + 34,50 \cdot \mathrm{NG}$
- $\beta_{0NG} = 8,6225 - 0,00108 \cdot 52,7 \approx 8,565$
- $\beta_{NG} = -0,00108 \cdot 34,50 + 0,303 \approx 0,266$

```
Call: lm(formula = log(PU) ~ NG, data = Zilli2020)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.56504    0.04257   201.2   <2e-16 ***
NG           0.26574    0.02607    10.2   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.2735 on 223 degrees of freedom
Multiple R-squared: 0.3179
F-statistic: 103.9 on 1 and 223 DF,  p-value: < 2.2e-16
  AIC   BIC
59.15 69.39
```

- Por isso a derivação de fatores através da regressão linear simples entre as variáveis explicativas e a variável independente não funciona!!!

Exemplo 3

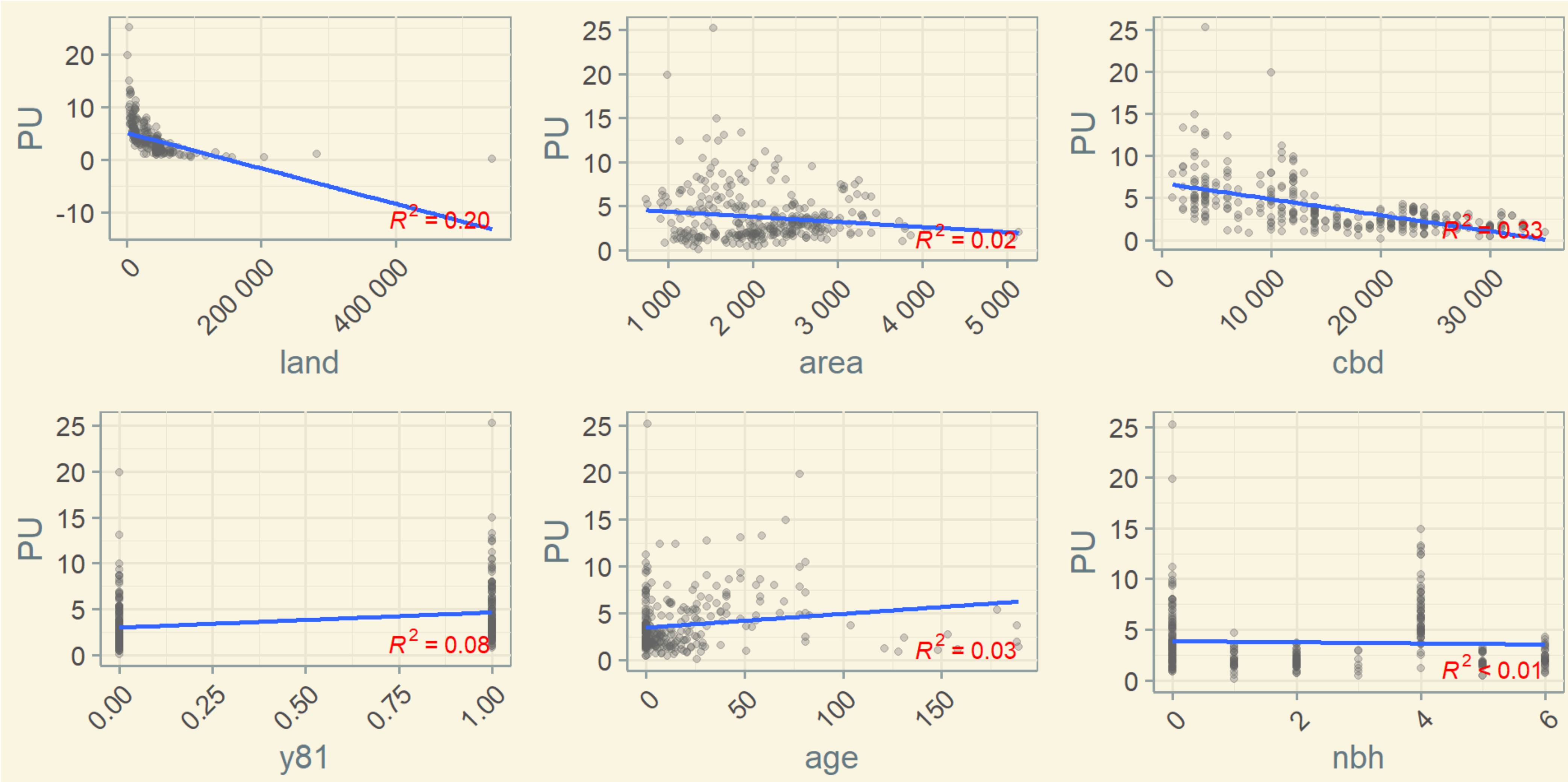


Dados

- Wooldridge (2019):

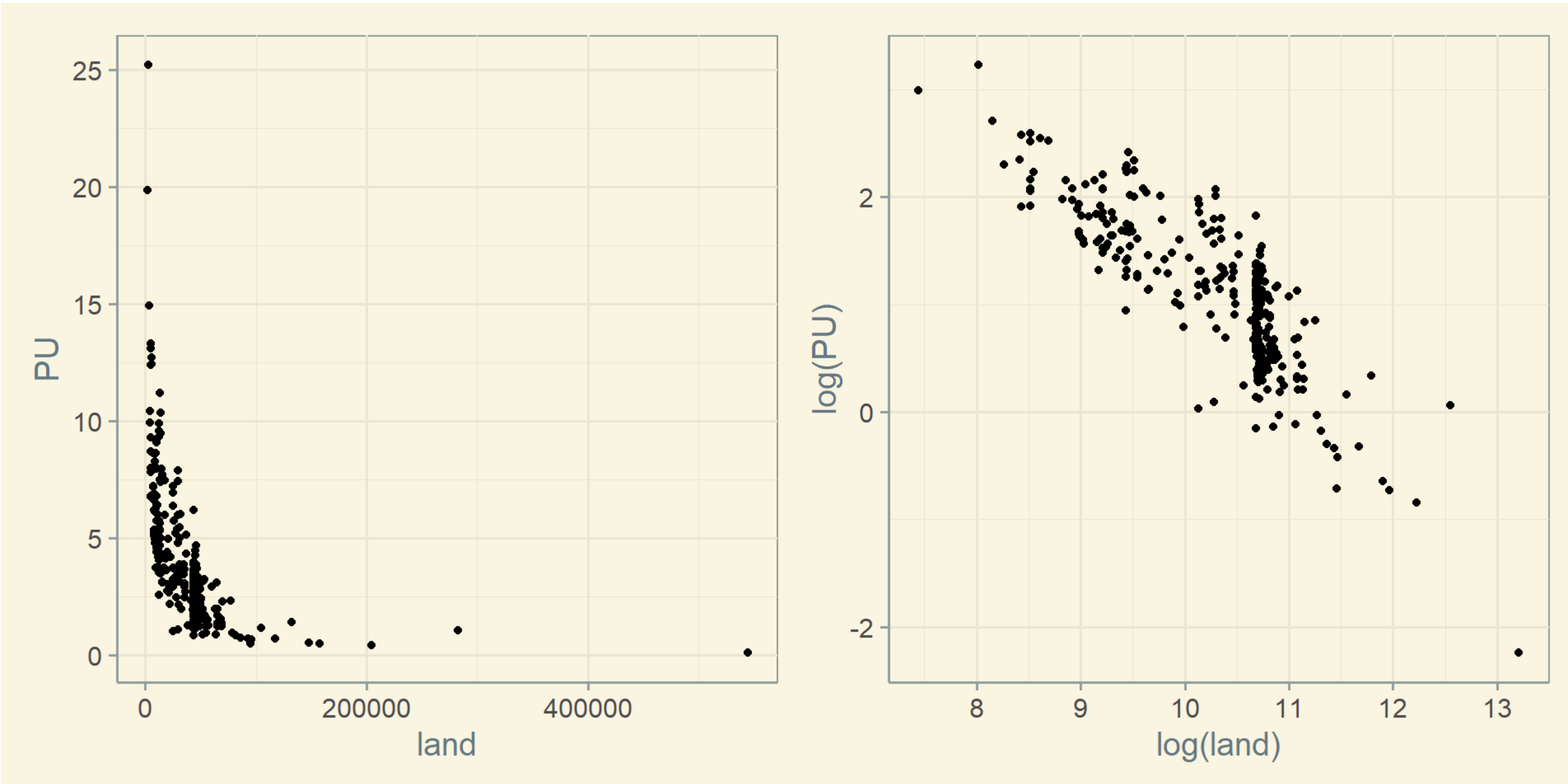
PU	price	area	land	year	y81	age	nbh	cbd	inst	rooms	baths	
13,11	60.000	1.660	4.578	1.978	0	48	4	3.000	1.000	7	1	10
4,78	40.000	2.612	8.370	1.978	0	83	4	4.000	1.000	6	2	11
6,80	34.000	1.144	5.000	1.978	0	58	4	4.000	1.000	6	1	11
6,39	63.900	1.136	10.000	1.978	0	11	4	4.000	1.000	5	1	11
4,40	44.000	1.868	10.000	1.978	0	48	4	4.000	2.000	5	1	12
4,84	46.000	1.780	9.500	1.978	0	78	4	3.000	2.000	6	3	10
5,15	56.000	1.700	10.878	1.978	0	22	4	4.000	2.000	6	2	11
9,95	38.500	1.556	3.870	1.978	0	78	4	3.000	2.000	6	2	10
8,64	60.500	1.642	7.000	1.978	0	42	4	3.000	2.000	8	2	10
6,92	55.000	1.443	7.950	1.978	0	41	4	3.000	2.000	5	2	11

Análise Exploratória

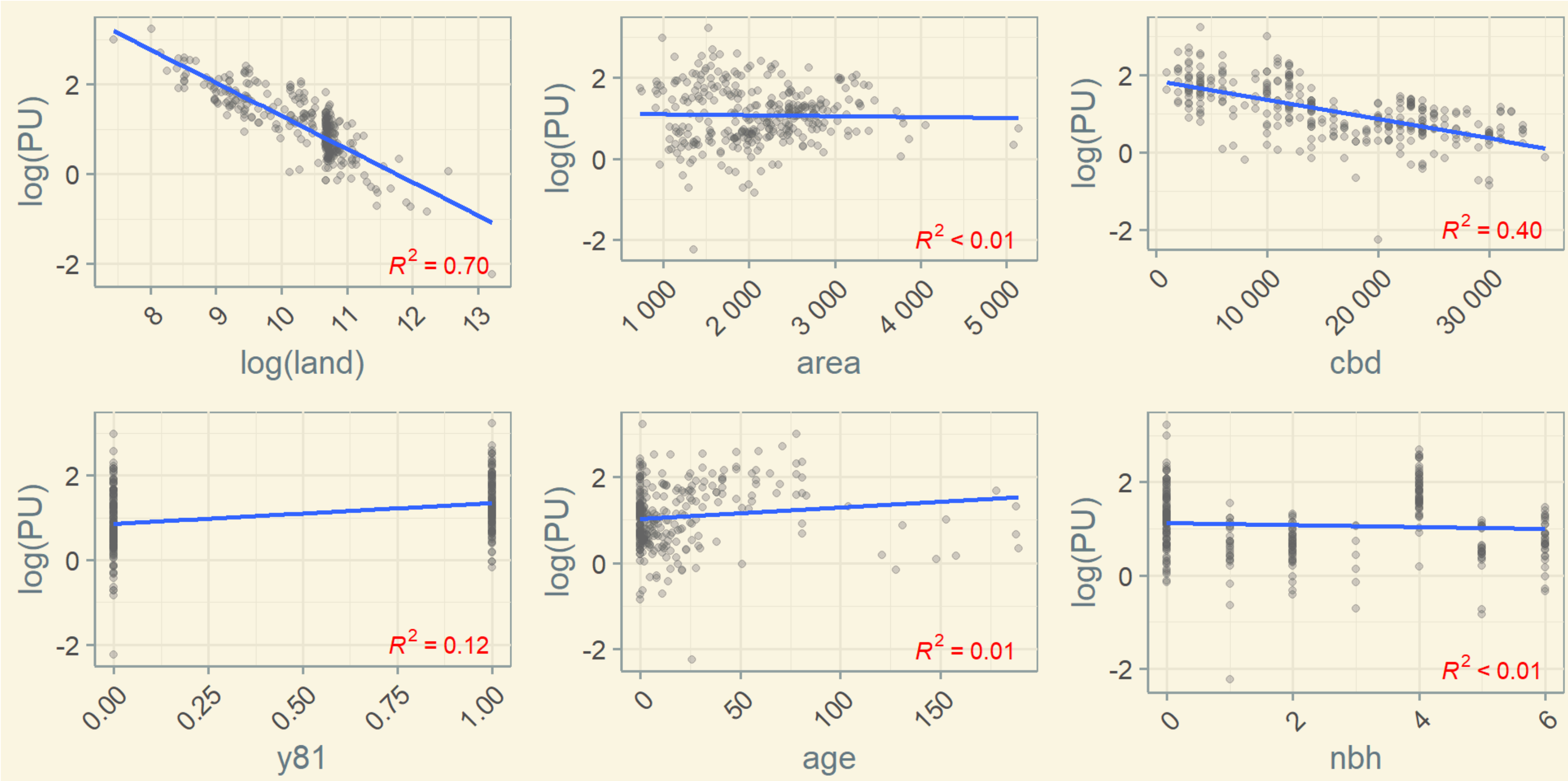


Análilse Exploratória

- Importante: a transformação $\ln()$ aproxima os dados:



Análise Exploratória



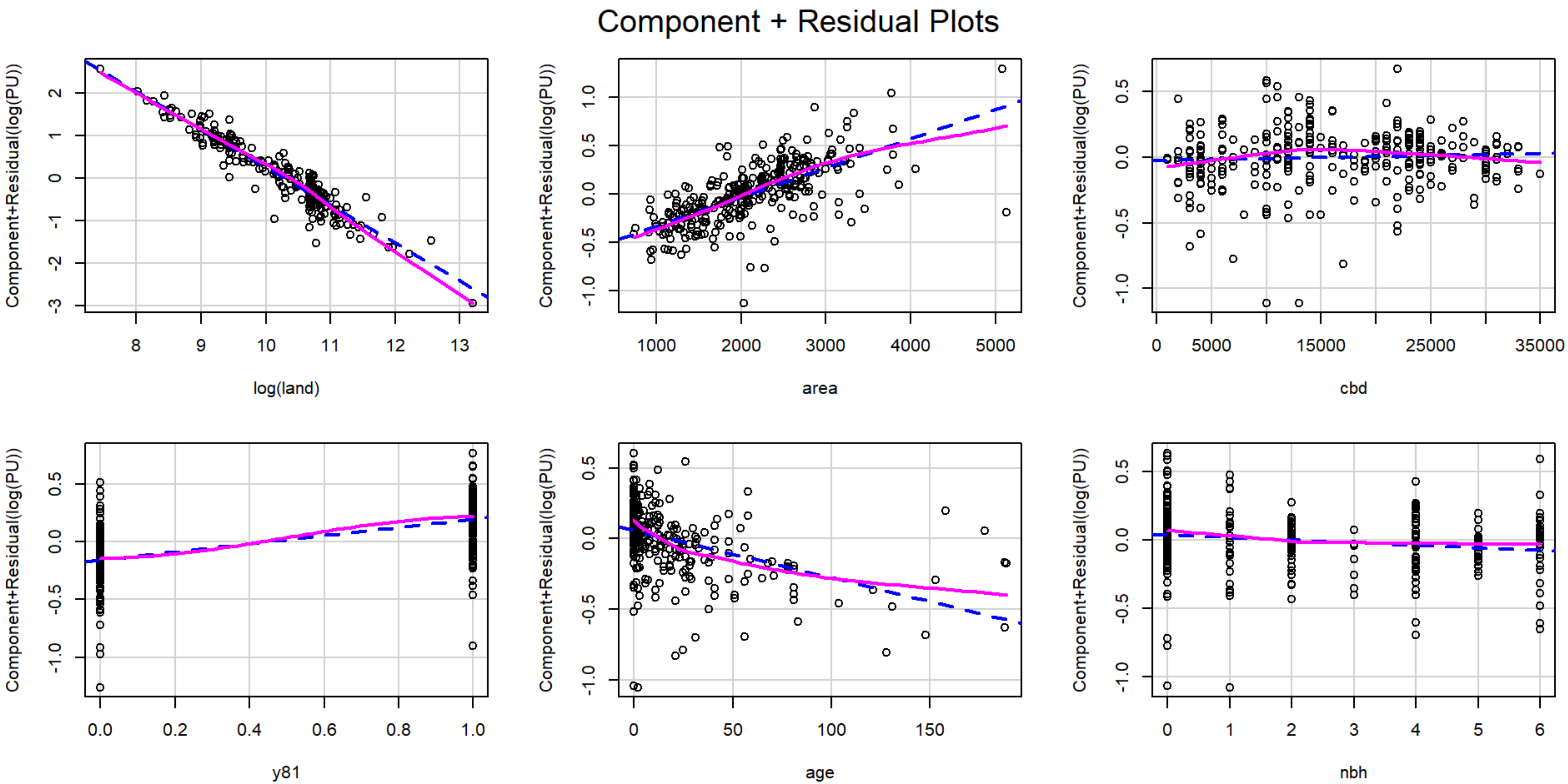
Ajuste de modelos

```
Call: lm(formula = log(PU) ~ log(land) + area + cbd + y81 + age + nbh, data =
         hprice3)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.548e+00  2.340e-01  40.804  < 2e-16 ***
log(land)   -8.914e-01  2.574e-02 -34.625  < 2e-16 ***
area         3.015e-04  2.039e-05  14.787  < 2e-16 ***
cbd          1.577e-06  2.294e-06   0.688  0.49221
y81          3.506e-01  2.654e-02  13.208  < 2e-16 ***
age         -3.376e-03  4.361e-04  -7.741 1.36e-13 ***
nbh         -1.964e-02  6.306e-03  -3.115  0.00201 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.2277 on 314 degrees of freedom
Multiple R-squared: 0.8983
F-statistic: 462.2 on 6 and 314 DF,  p-value: < 2.2e-16
   AIC    BIC
-30.23  -0.06
```

Diagnóstico do modelo

- Para a estimação de fatores é importante atentar para o bom ajuste do modelo
  - Com um modelo mal ajustado não será possível derivar fatores precisos
  - Algumas ferramentas possibilitam a verificação da linearidade do modelo, por exemplo

Diagnóstico do modelo



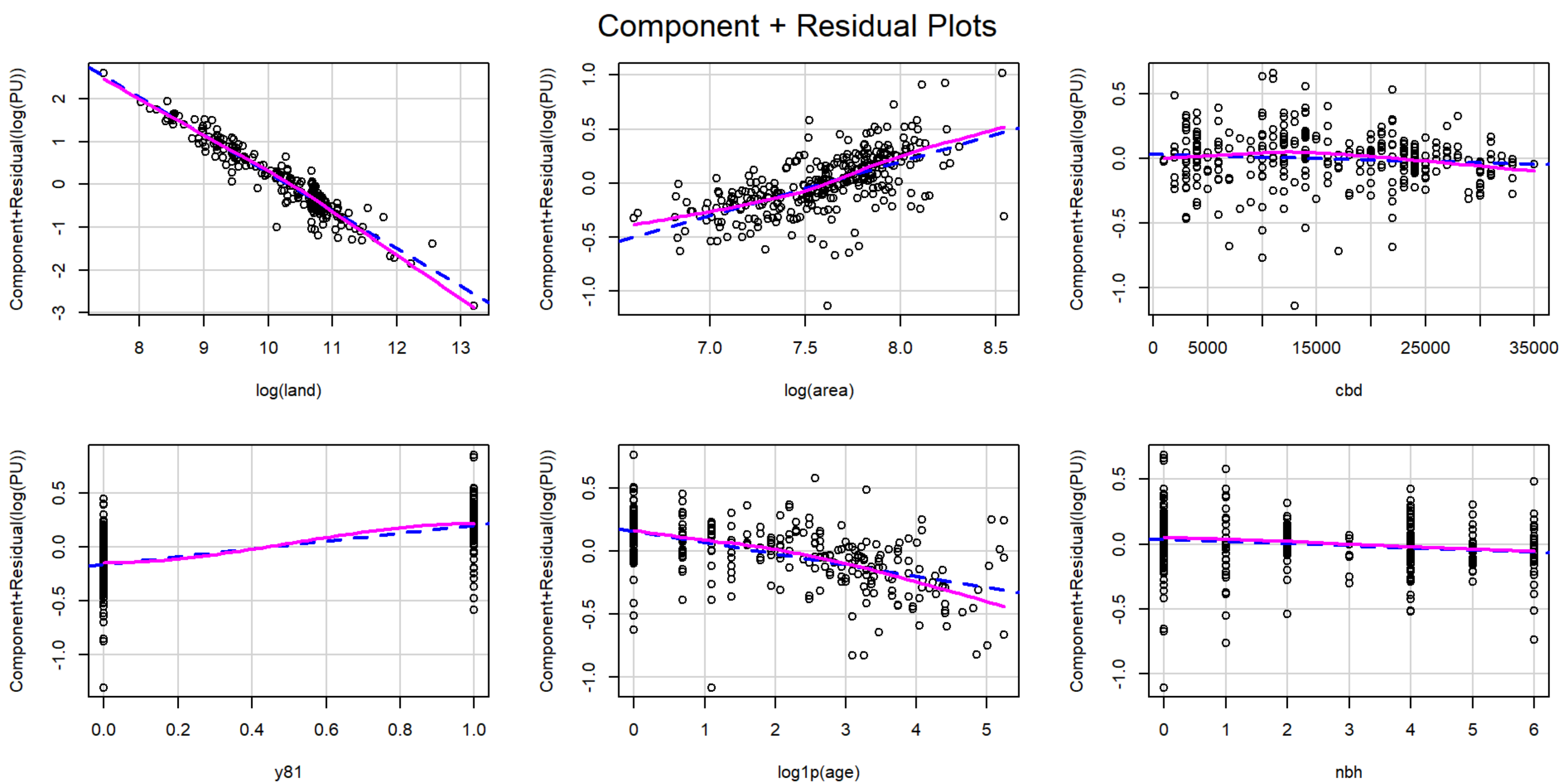
- A variável *Area* e a variável *Age* parecem pedir transformação!

Novo modelo

```
Call: lm(formula = log(PU) ~ log(land) + log(area) + cbd + y81 + log1p(age) +
         nbh, data = hprice3)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.456e+00  3.258e-01  19.813  < 2e-16 ***
log(land)   -8.821e-01  2.404e-02 -36.701  < 2e-16 ***
log(area)    4.970e-01  4.066e-02  12.224  < 2e-16 ***
cbd         -2.097e-06  2.202e-06  -0.952  0.34170
y81          3.599e-01  2.478e-02  14.525  < 2e-16 ***
log1p(age)  -8.930e-02  9.367e-03  -9.533  < 2e-16 ***
nbh         -1.707e-02  5.952e-03  -2.868  0.00441 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.2144 on 314 degrees of freedom
Multiple R-squared: 0.9098
F-statistic:   528 on 6 and 314 DF,  p-value: < 2.2e-16
   AIC    BIC
-68.83 -38.66
```

- A variável *cbd* não se mostrou estatisticamente significante!

Diagnóstico do novo modelo

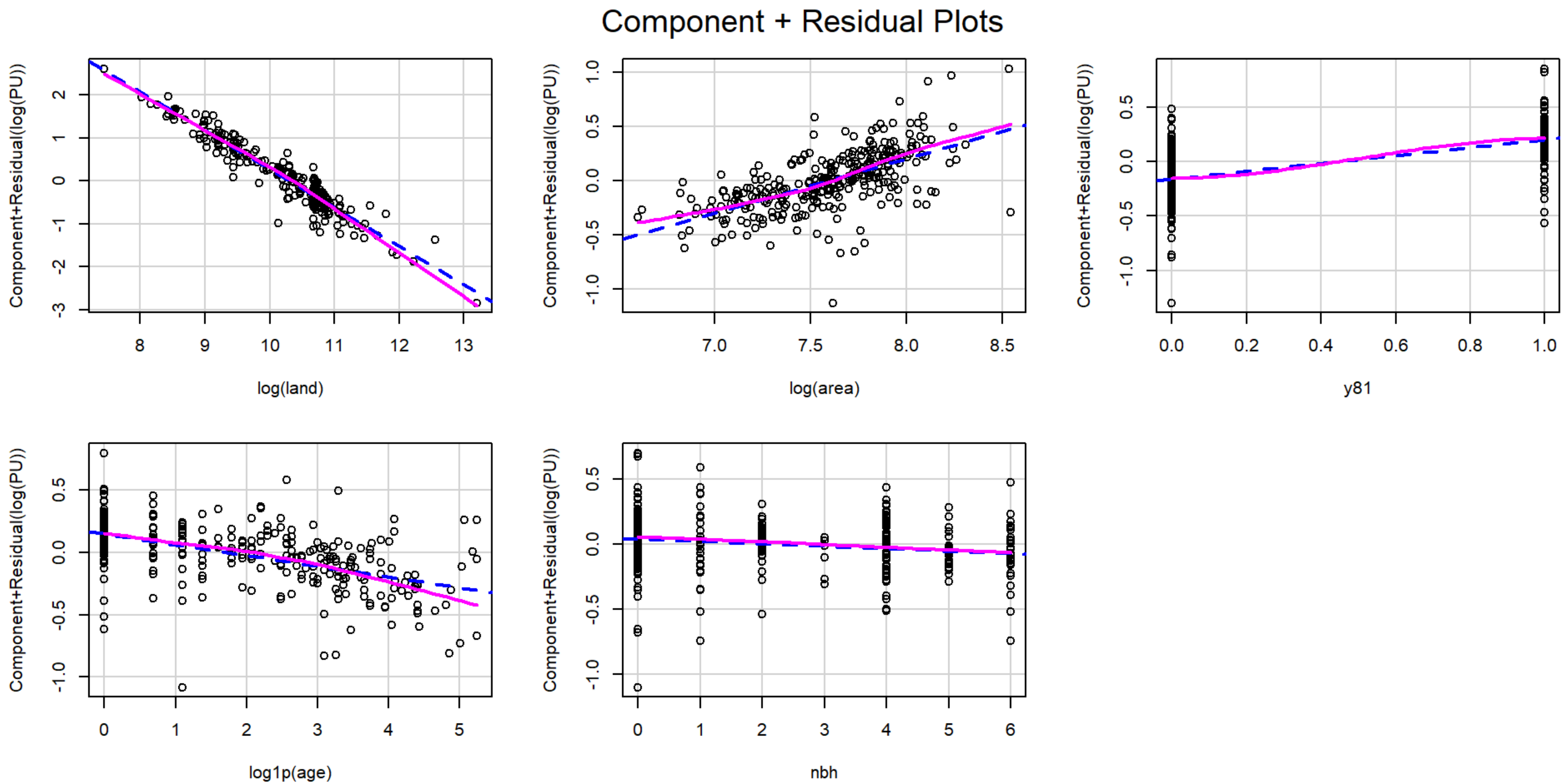


- As variáveis mais importantes são: *land*, *area*, *age* e *y81*
  - *cbd* e *nbh* são duas variáveis pouco importantes!

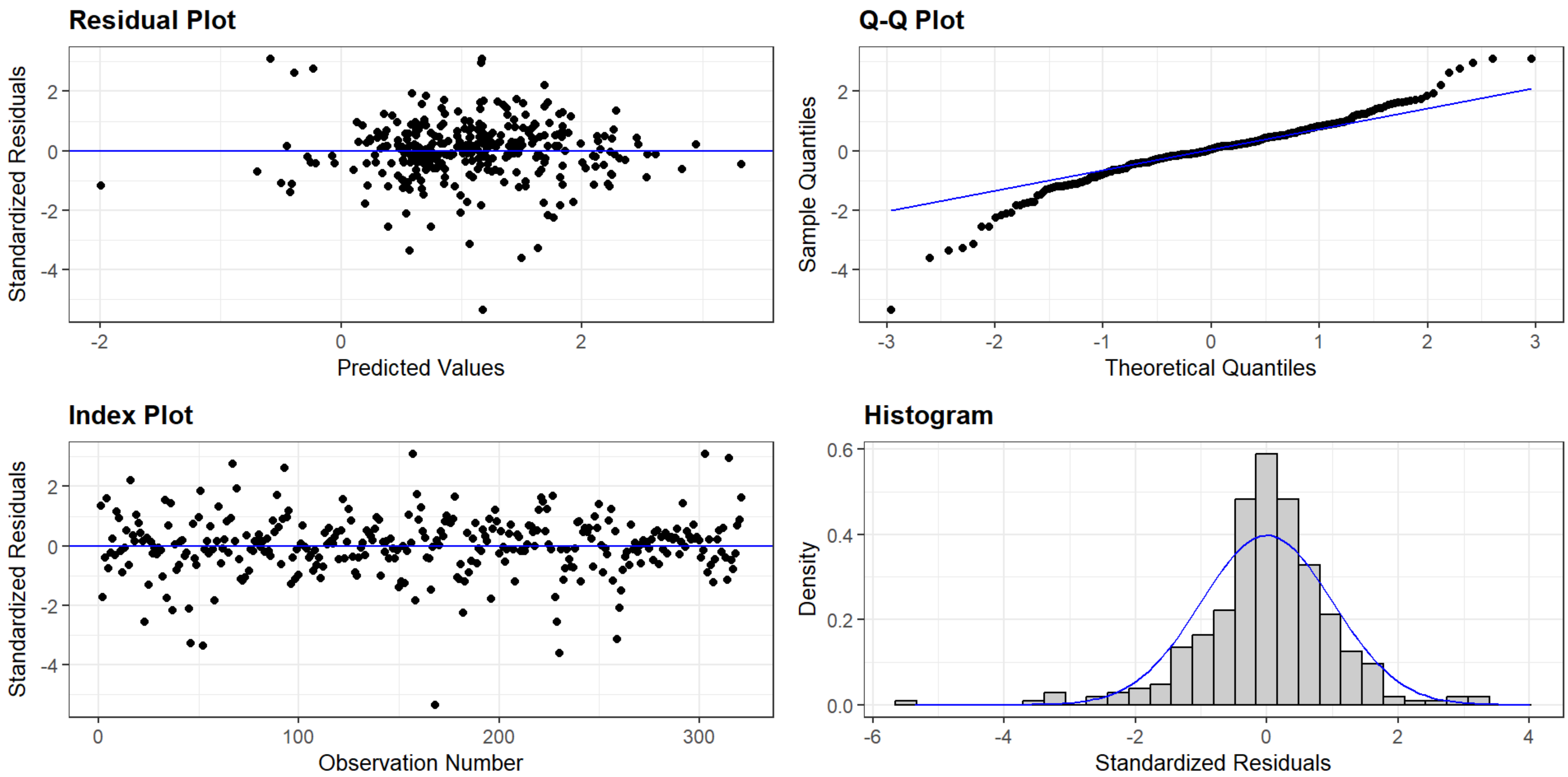
Modelo Final

```
Call: lm(formula = log(PU) ~ log(land) + log(area) + y81 + log1p(age) + nbh,
         data = hprice3)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.528721   0.316575  20.623  < 2e-16 ***
log(land)   -0.897794   0.017509 -51.275  < 2e-16 ***
log(area)    0.504073   0.039980  12.608  < 2e-16 ***
y81          0.361585   0.024707  14.635  < 2e-16 ***
log1p(age)  -0.086262   0.008807  -9.795  < 2e-16 ***
nbh         -0.019047   0.005578  -3.415 0.000722 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.2143 on 315 degrees of freedom
Multiple R-squared: 0.9096
F-statistic: 633.6 on 5 and 315 DF,  p-value: < 2.2e-16
  AIC   BIC
-69.9 -43.5
```

Diagnóstico do Modelo Final



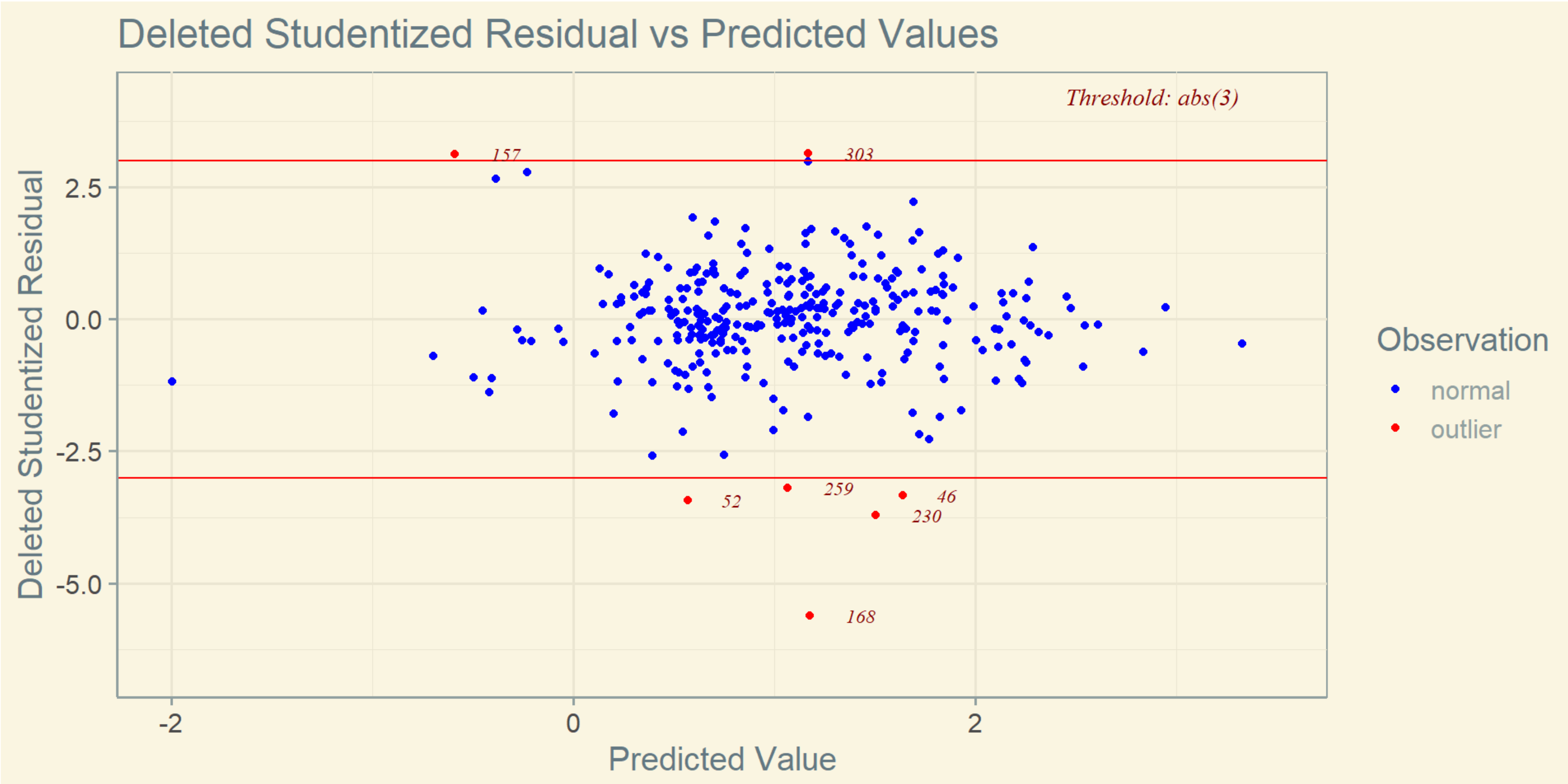
Diagnóstico do Modelo Final



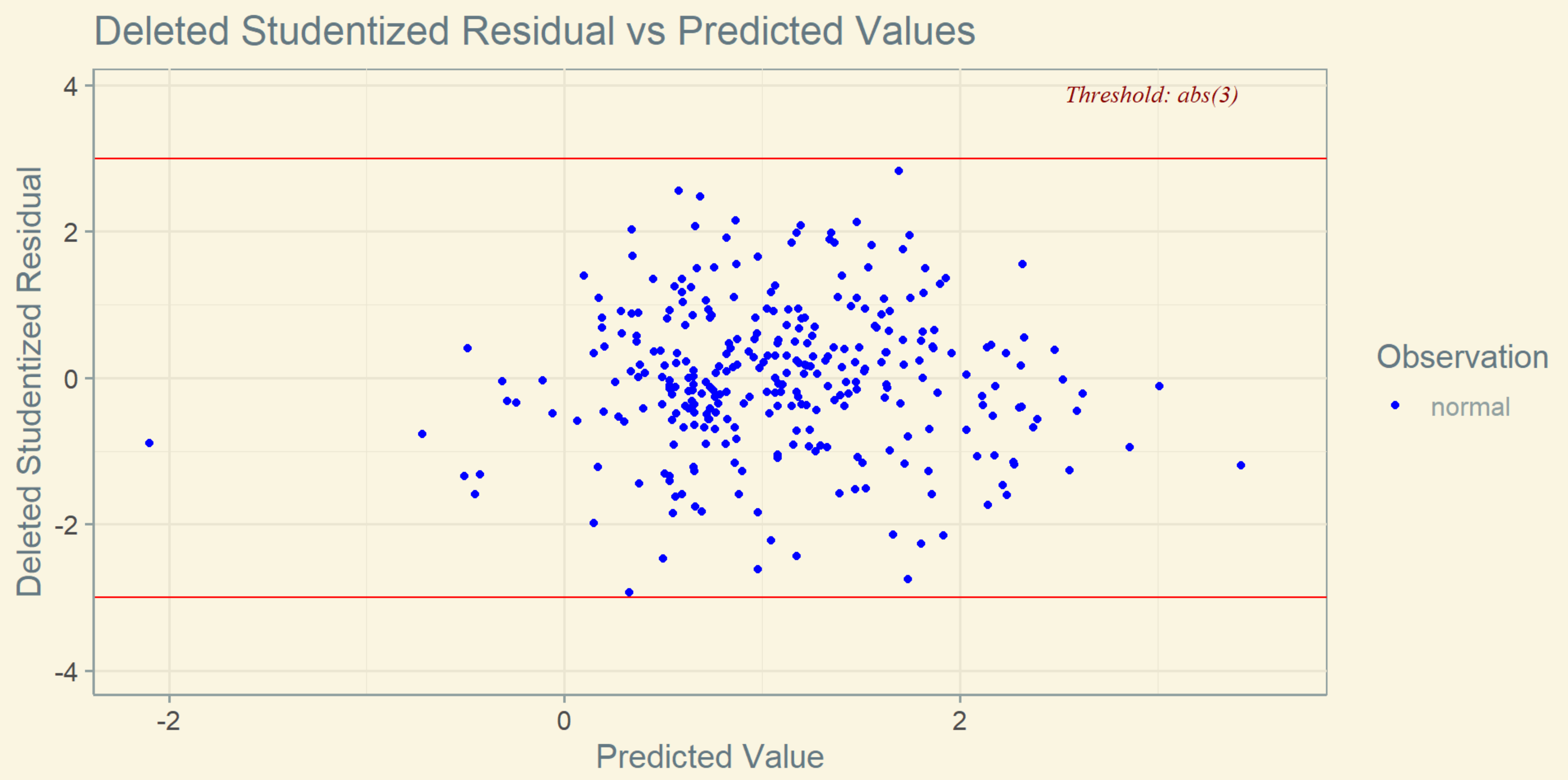
Aparte: Resíduos *Jackknife*

- Resíduos *jackknife* são resíduos padronizados que são computados ao realizar um processo de reamostragem conhecido como *jackknife*!
- Grosso modo, o processo funciona da seguinte forma:
  - Retira-se um dado da amostra
  - Ajusta-se um modelo
  - Computa-se o valor previsto com este modelo para o dado excluído da amostra
  - Calcula-se o seu resíduo padronizado deste dado
  - Prossegue-se retirando um dado de cada vez até que todos os dados tenham seus resíduos calculados
- Também são conhecidos como resíduos *studentizados* (t_i)
- Um bom critério para detecção de *outliers*: $|t_i| > 3,0$

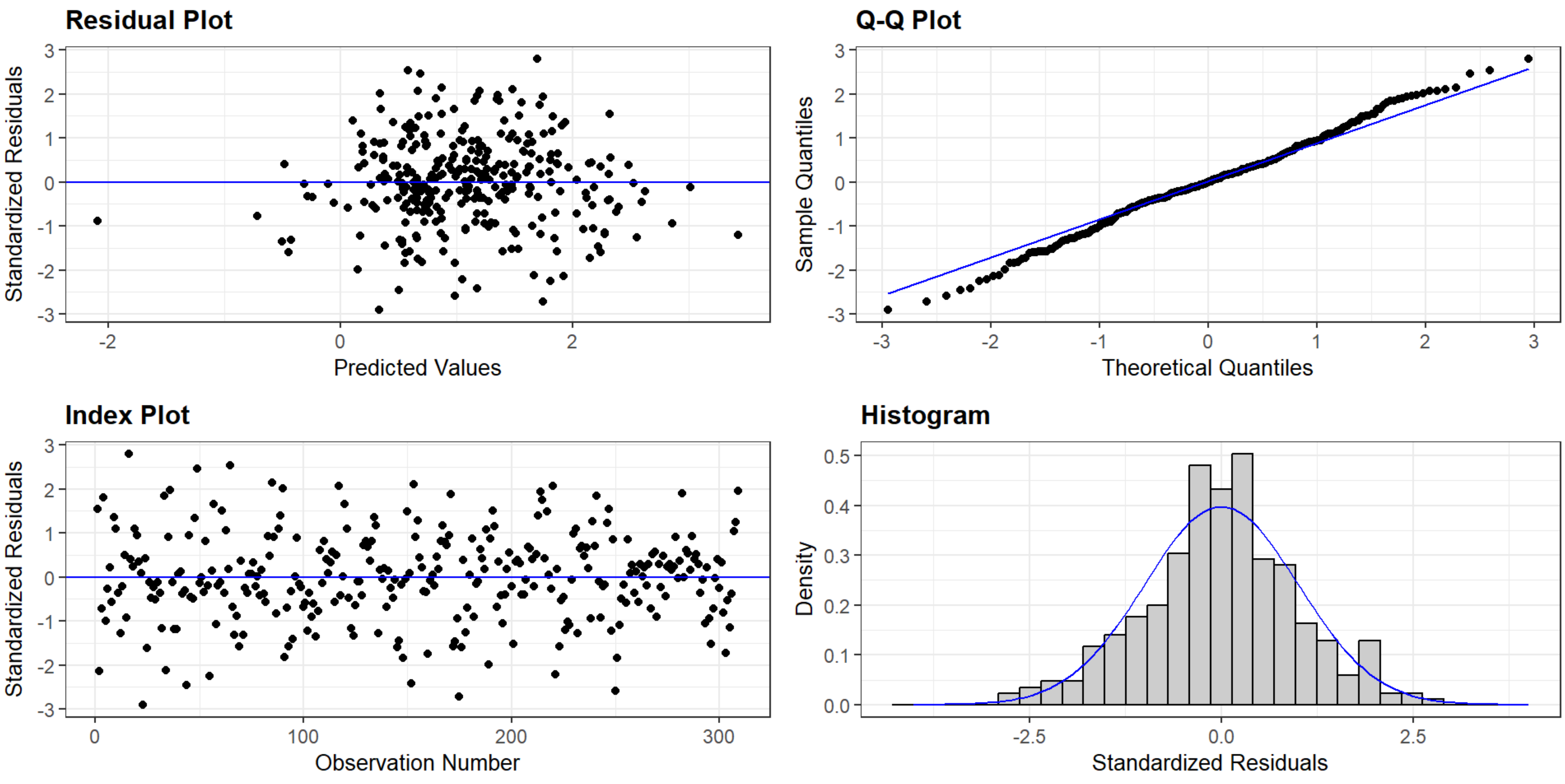
Diagnóstico do Modelo Final



Reajustamento do Modelo Final



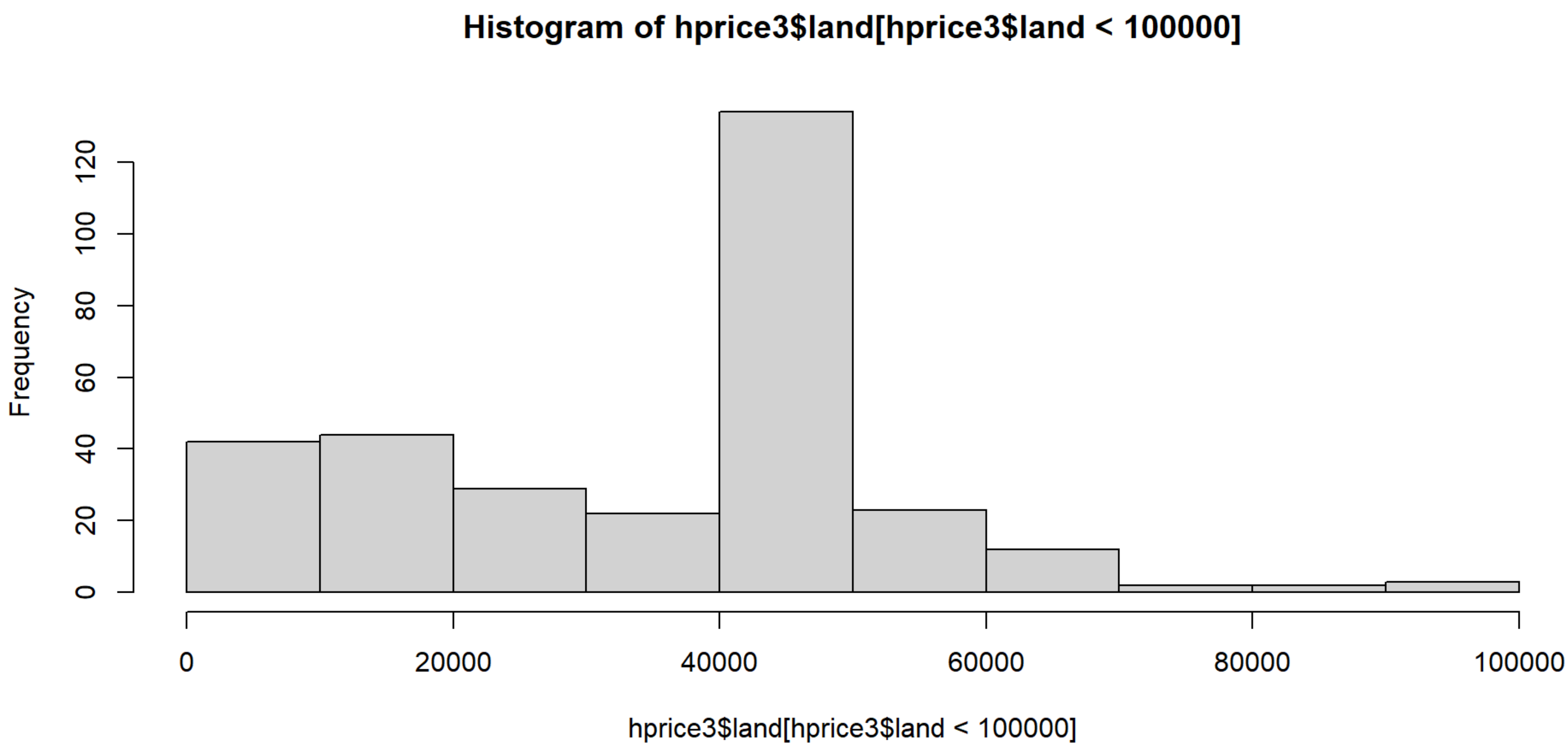
Painel de Resíduos do Modelo Final



Estatísticas do Modelo Final

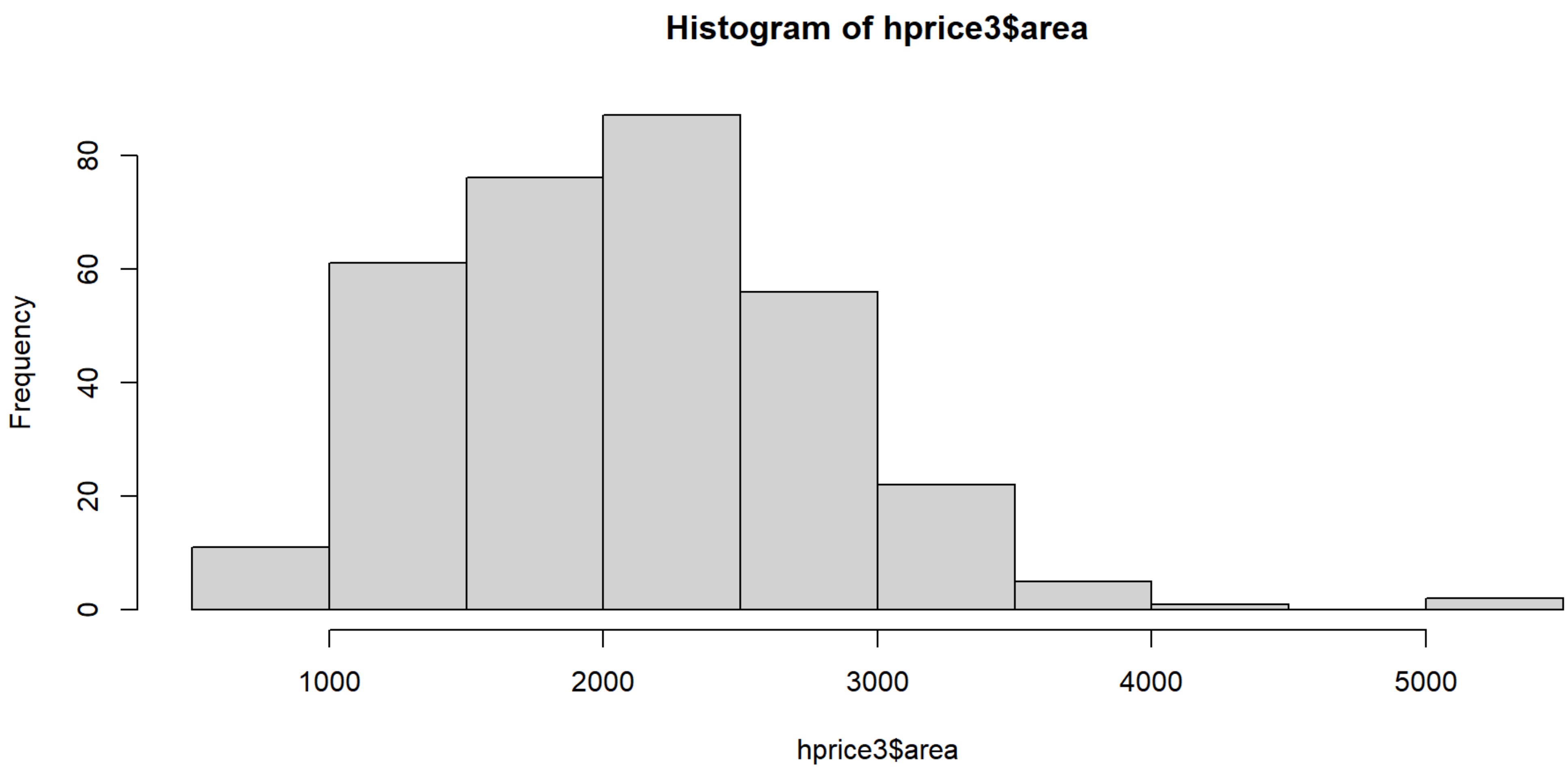
```
Call: lm(formula = log(PU) ~ log(land) + log(area) + y81 + log1p(age) + nbh,
         data = hprice3, subset = -c(37, 46, 52, 67, 93, 157, 168, 229, 230, 259, 303,
         315))

Coefficients:
             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  7.058832   0.266793  26.458   < 2e-16 ***
log(land)   -0.928501   0.014185 -65.455   < 2e-16 ***
log(area)    0.481479   0.032656  14.744   < 2e-16 ***
y81          0.354456   0.019689  18.003   < 2e-16 ***
log1p(age)  -0.104942   0.007338 -14.302   < 2e-16 ***
nbh         -0.019478   0.004465  -4.363 0.0000176 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.1676 on 303 degrees of freedom
Multiple R-squared: 0.9442
F-statistic:  1024 on 5 and 303 DF,  p-value: < 2.2e-16
    AIC     BIC
-218.92 -192.79
```

Centralização do Modelo Final



- Moda: 43.560,00 sqft $\approx 1acre \approx 4046,85m^2$

Centralização do Modelo Final



- Moda: 2.464 sqft $\approx 230m^2$

Outras variáveis

- y81: 0 (não)
- age: 0 (novo)
- nbh: 0

Modelo Centralizado

```
Call: lm(formula = log(PU) ~ log(land/43560) + log(area/2464) + y81 +
         log1p(age) + nbh, data = hprice3, subset = -c(37, 46, 52, 67, 93, 157, 168,
         229, 230, 259, 303, 315))

Coefficients:
                 Estimate Std. Error t value  Pr(>|t|)
(Intercept)      0.900808   0.020276  44.427   < 2e-16 ***
log(land/43560) -0.928501   0.014185 -65.455   < 2e-16 ***
log(area/2464)   0.481479   0.032656  14.744   < 2e-16 ***
y81              0.354456   0.019689  18.003   < 2e-16 ***
log1p(age)      -0.104942   0.007338 -14.302   < 2e-16 ***
nbh             -0.019478   0.004465  -4.363 0.0000176 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 0.1676 on 303 degrees of freedom
Multiple R-squared: 0.9442
F-statistic:  1024 on 5 and 303 DF,  p-value: < 2.2e-16
    AIC     BIC
-218.92 -192.79
```

Derivação de Fatores

- $PU_{Hom} = \exp(\hat{\beta}_0^*) = \exp(0,90) \approx 2,46$
- $F_S = \left(\frac{43.560}{S_i}\right)^{0,93}$
- $F_A = \left(\frac{A_i}{2.464}\right)^{0,48}$
- $F_Y = \begin{cases} 1 & \text{se } y_{81i} = 0; \\ \exp(.35) = 1,425 & \text{se } y_{81i} = 1; \end{cases}$
- $F_I = (1 + Age_i)^{-0,105}$
- $F_N = \exp(-0,02)^{nbh_i} = 0,98^{nbh}$

Aplicação

- Avaliar por fatores:
  - $S = 360m^2; A = 100m^2, y_{81} = 1, age = 0, nbh = 6$
  - Obs.: $1m^2 = 10,764\, sqft$
- Solução pelo modelo:
- $$\begin{aligned}\ln(\widehat{PU}) = 0,90 - 0,93\ln(S/43.560) + 0,48 \cdot \ln(A/2.464) + \\ 0,3545 \cdot y_{81} - 0,105\ln(1 + age) - 0,02 \cdot nbh\end{aligned}$$
- $$\begin{aligned}\ln(\widehat{PU}) = 0,90 - 0,93\ln(3.875/43.560) + 0,48 \cdot \ln(1.076,4/2.464) + \\ 0,3545 \cdot 1,0 - 0,105\ln(1 + 0) - 0,02 \cdot 6\end{aligned}$$
- $$\ln(\widehat{PU}) = 0,90 + 2,25 - 0,40 + 0,3545 - 0 - 0,12 = 2,9845$$
- $\widehat{PU} = \exp(2,9845) \approx$ 19,78 (US$/sqft)

Solução por fatores

- $F_S = \left(\frac{43.560}{S_i}\right)^{0{,}93} = \left(\frac{43.560}{3.875}\right)^{0{,}93} = 9,49$
- $F_A = \left(\frac{A_i}{2.464}\right)^{0{,}48} = \left(\frac{1.076{,}4}{2.464}\right)^{0{,}48} = 0,67$
- $F_Y = 1,425$
- $F_I = (1 + Age_i)^{-0{,}105} = (1 + 0)^{-0{,}105} = 1,0$
- $F_N = 0,98^{nbh} = 0,98^6 = 0,886$
- $\widehat{PU}_i = 2,46 \cdot F_{Si} \cdot F_{Ai} \cdot F_{Y_i} \cdot F_{Ii} \cdot F_{Ni}$
- $\widehat{PU}_i = 2,46 \cdot 9,49 \cdot 0,67 \cdot 1,425 \cdot 1,0 \cdot 0,886$
- $\widehat{PU}_i = 2,46 \cdot 8,03 \approx 19,75$ (US$/sqft)

Análise

Var.	n_missing	%	Mean	SD	P0	P25	P50	P75	P100	Hist.
PU	0	1	3,75	2,97	0,11	1,86	2,92	4,78	25,22	
price	0	1	96.100,66	43.223,73	26.000,00	65.000,00	85.900,00	120.000,00	300.000,00	
area	0	1	2.106,73	694,96	735,00	1.560,00	2.056,00	2.544,00	5.136,00	
land	0	1	39.629,89	39.514,39	1.710,00	16.935,00	43.560,00	46.100,00	544.500,00	
year	0	1	1.979,33	1,49	1.978,00	1.978,00	1.978,00	1.981,00	1.981,00	
y81	0	1	0,44	0,50	0,00	0,00	0,00	1,00	1,00	
age	0	1	18,01	32,57	0,00	0,00	4,00	22,00	189,00	
nbh	0	1	2,21	2,16	0,00	0,00	2,00	4,00	6,00	
cbd	0	1	15.822,43	8.967,11	1.000,00	9.000,00	14.000,00	23.000,00	35.000,00	
inst	0	1	16.442,37	9.033,13	1.000,00	9.000,00	16.000,00	24.000,00	34.000,00	
rooms	0	1	6,59	0,90	4,00	6,00	7,00	7,00	10,00	
baths	0	1	2,34	0,77	1,00	2,00	2,00	3,00	4,00	
dist	0	1	20.715,58	8.508,18	5.000,00	13.400,00	19.900,00	27.200,00	40.000,00	

- Range da Var. *land*: 1.710, 544.500 sqft (159 a 50.585 m^2)
- Range da Var. *area*: 735, 5.136 sqft (68 a 477 m^2)
- Range da Var. *age*: 0, 189 anos

Referências

Cerino, Rocío Mariel, Juan Pablo Carranza, Mario Andrés Piumetto, María Emilia Bullano, Federico Monzani, e Mariano Augusto Córdoba. 2020. «Homogeneización del valor de la tierra urbana mediante técnicas de econometría espacial en valuaciones masivas automatizadas». *14º COBRAC*. https://rdu.unc.edu.ar/handle/11086/28446.

Droubi, Luiz Fernando Palin, e Lutemberg de Araújo Florencio. 2024. «Fatores de Homogeneização: Derivação e uso racional». *XXXVIII Congresso Pan-Americano UPAV – CHILE 2024*. https://www.upavchile2024.com/.

Droubi, Luiz Fernando Palin, Carlos Augusto Zilli, e Norberto Hochheim. 2021. «Avaliação de Imóveis à partir de dados de oferta e transação». *XXI COBREAP*.

Lima, Gilson Pereira de Andrade. 2006. «Homogeneização por Fatores na Forma Aditiva, Multiplicativa ou Mista? Imposição Normativa ou Resposta do Mercado?» *XIII COBREAP*.

Oliveira, Antônio Augusto Ferreira de, Sandro Ricardo Vasconcelos Bandeira, e Victor Jucá Távora. 2024. «Mass appraisal of urban land with homogenization factors: a spatial models-based approach». *Revista Valorem* 1 (1): 16–32. https://revistavalorem.com/index.php/home/article/view/23.

Trivelloni, Carlos Alberto Peruzzo. 2005. «Método para determinação do valor da localização com uso de técnicas inferenciais e geoestatísticas na avaliação em massa de imóveis». Tese de Doutorado, Florianópolis, SC: Universidade Federal de Santa Catarina.

Wooldridge, Jeffrey M. 2019. *Introductory Econometrics: A Modern Approach*. 7.ª ed. Thomson South-Western.

Zeni, André Maciel. 2024. *Engenharia de Avaliações: Metodologia Científica - Processo Comparativo*. Feedback.

Zilli, Carlos Augusto. 2020. «Regressão geograficamente ponderada aplicada na avaliação em massa de imóveis urbanos.» Dissertação de mestrado. Florianópolis, SC: Universidade Federal de Santa Catarina.