

DESAFIOS NA AVALIAÇÃO DE IMÓVEIS POR REGRESSÃO LINEAR COM PEQUENAS AMOSTRAS

**Autores: Rubens Dantas; Érico Lisboa,
Maria Marques**

João Pessoa, 20 de novembro 2025

DESENVOLVIMENTO

- LIMITAÇÕES DO MODELO
- ENQUADRAMENTO NORMATIVO
- DESAFIOS PARA CONTORNAR PROBLEMAS
Bootstrap, Validação cruzada, Regressão Robusta –
Método de Huber (RLM), Regressão RANSAC (Random
Sample Consensus), Inferência Baysiana,
- CONSIDERAÇÕES FINAIS

LIMITAÇÕES DO MODELO

- ALTA VARIABILIDADE NOS COEFICIENTES
- OVERFITTING x UNDERFITTING
- AMPLO INTERVALO DE CONFIANÇA
- ALTA SIGNIFICÂNCIA (teste-T)
- AUMENTO DA VARIÂNCIA

CASO: INTERIOR DO ESTADO DO PARÁ

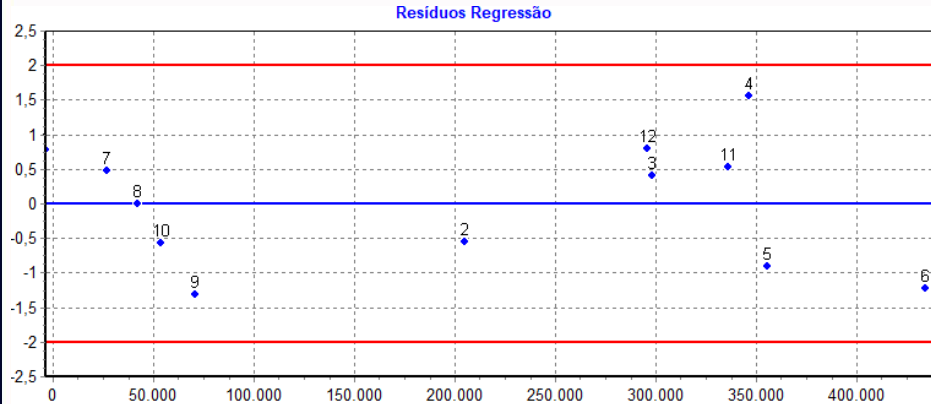
X1	X2	Y
30	250	R\$ 18.000,00
275	630	R\$ 190.000,00
450	720	R\$ 310.000,00
595	700	R\$ 390.000,00
680	625	R\$ 330.000,00
800	735	R\$ 400.000,00
70	300	R\$ 40.000,00
70	350	R\$ 42.000,00
65	450	R\$ 34.500,00
60	400	R\$ 38.140,00
555	715	R\$ 351.000,00
500	650	R\$ 318.000,00

Coeficientes	
Correlação	1 - 0,9874 / 0,9874
Determinação	1 - 0,9749 / 0,9749
R2 Ajustado	1 - 0,9693 / 0,9693
Testes de Hipóteses	
F Calculado	174,7
Significância Mod	0,00%(0,00000)
Grau Fundamenta	Grau III

Valor total =
 -91746,98106
 +374,6959389 * Area total
 +307,6805557 * Distancia ao polo valorizante

R² é muito elevado com poucos dados: O modelo se ajustou muito aos pontos específicos da amostra, em vez de captar o comportamento geral da população. Pode ser um **sinal clássico de "overfitting".**

CASO: INTERIOR DO ESTADO DO PARÁ



Coeficientes

Correlação 1 - 0,9874 / 0,9874

Determinação 1 - 0,9749 / 0,9749

R2 Ajustado 1 - 0,9693 / 0,9693

Testes de Hipóteses

F Calculado 174,7

Significância Mod 0,00%(0,00000)

Grau Fundamenta Grau III

Heterocedasticidade significativa

($p < 0.01$ no Breusch–Pagan) Isso indicou:

- o modelo não generaliza bem,
- os erros variam conforme os valores previstos,
- há problemas estruturais que aumentam risco de overfitting.

Valor total =

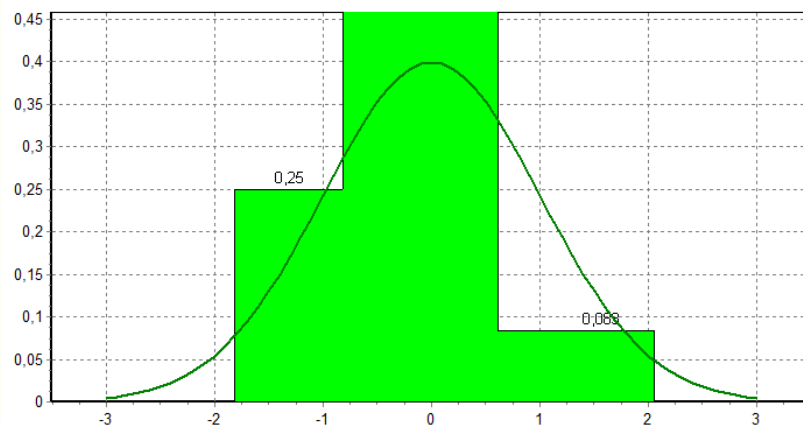
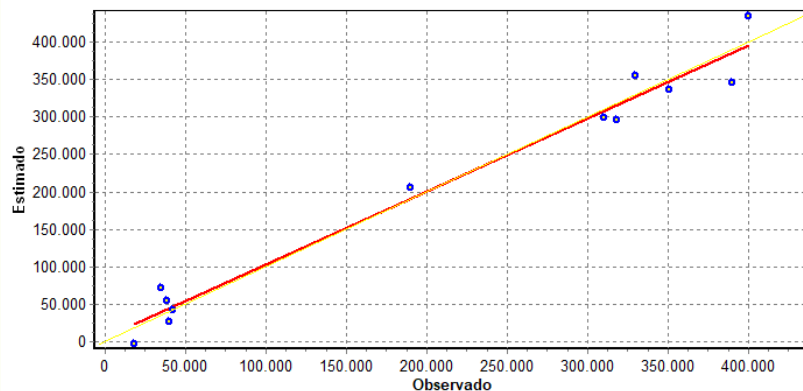
-91746,98106

+374,6959389 * Area total

+307,6805557 * Distancia ao polo valorizante

O $F = 174,7$ é **extremamente elevado** considerando: pequena amostra, modelo simples. Modelos superajustados tendem a apresentar estatísticas infladas.

CASO: INTERIOR DO ESTADO DO PARÁ



Normalidade dos Resíduos

-1 e +1 desvios padrões

75% [68%]

-1,64 e +1,64 desvios padrões

100% [90%]

-1,96 e +1,96 desvios padrões

100% [95%]

Diversos

Desvio Padrão

27894

Outliers do Modelo

0 (0,00%)

Método de Cálculo

Geral

Variáveis	Transf.	Elasticidade	t Calculado	Sig.(%)
Area total	x	14,06%	5,78	0,03%
Distancia ao polo valorizante	x	7,27%	3,04	1,40%
Valor total	y		-2,45	3,69%

CASO: INTERIOR DO ESTADO DO PARÁ

X1	X2	Y
30	250	R\$ 18.000,00
275	630	R\$ 190.000,00
450	720	R\$ 310.000,00
595	700	R\$ 390.000,00
680	625	R\$ 330.000,00
800	735	R\$ 400.000,00
70	300	R\$ 40.000,00
70	350	R\$ 42.000,00
65	450	R\$ 34.500,00
60	400	R\$ 38.140,00
555	715	R\$ 351.000,00
500	650	R\$ 318.000,00

Coeficientes

Correlação 1 - 0,9931 / 0,9939

Determinação 1 - 0,9863 / 0,9879

R2 Ajustado 1 - 0,9832 / 0,9852

Testes de Hipóteses

F Calculado 323,7

Significância Modelo 0,00%(0,00000)

Grau Fundamentação ■ Grau III

Valor total = (
-385,4159204
+134,0835696 * ln (Area total)
+0,0002409052179 * Distancia ao polo valorizante²)²

R² é muito elevado com poucos dados: O modelo **se ajustou muito aos pontos específicos da amostra**, em vez de captar o comportamento geral da população. Pode ser um **sinal clássico de “overfitting”**.

CASO: INTERIOR DO ESTADO DO PARÁ



Coeficientes

Correlação	1 - 0,9931 / 0,9939
Determinação	1 - 0,9863 / 0,9879
R2 Ajustado	1 - 0,9832 / 0,9852

Testes de Hipóteses

F Calculado	323,7
Significância Modelo	0,00%(0,00000)
Grau Fundamentação	 Grau III

Heterocedasticidade significativa

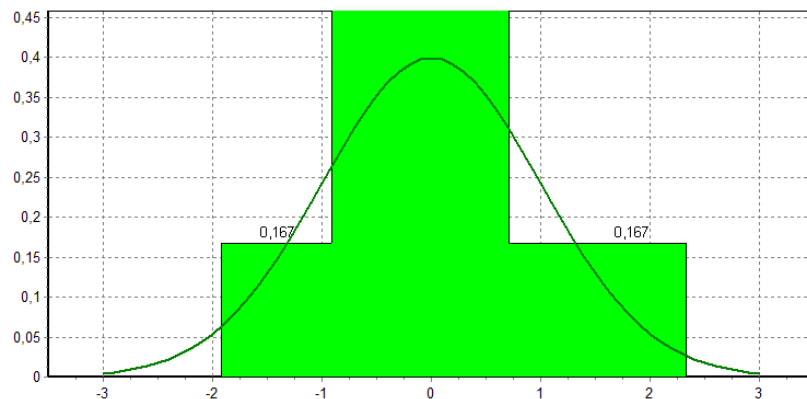
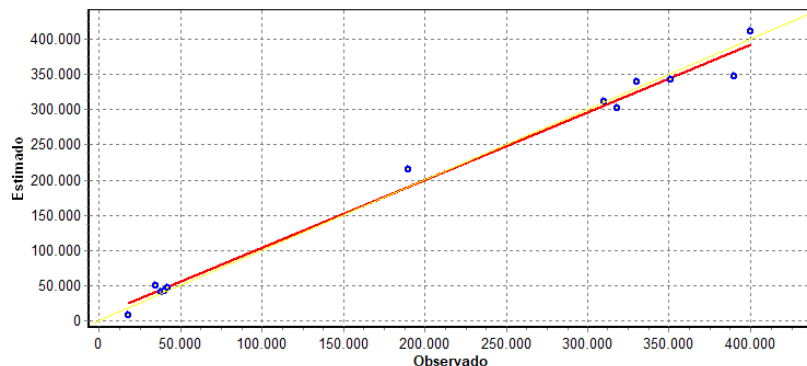
($p < 0.01$ no Breusch–Pagan) Isso indicou:

- o modelo não generaliza bem,
- os erros variam conforme os valores previstos,
- há problemas estruturais que aumentam risco de overfitting.

Valor total = (
-385,4159204
+134,0835696 * ln (Area total)
+0,0002409052179 * Distancia ao polo valorizante²)²

O $F = 323,7$ é **extremamente elevado** considerando: pequena amostra, modelo simples. Modelos superajustados tendem a apresentar estatísticas infladas.

CASO: INTERIOR DO ESTADO DO PARÁ



Normalidade dos Resíduos

-1 e +1 desvios padrões	66% [68%]
-1,64 e +1,64 desvios pa	91% [90%]
-1,96 e +1,96 desvios pa	100% [95%]

Diversos

Desvio Padrão	26,483
Outliers do Modelo	0 (0,00%)
Método de Cálculo	Geral

Variáveis	Transf.	Elastic...	t Calcul...	Sig. (%)
Area total	$\ln(x)$	11,81	5,97	0,02
Distancia ao ...	x^2	5,73	1,65	13,37
Valor total	$y^{1/2}$		-5,06	0,07

CASO: INTERIOR DO ESTADO DO PARÁ

Regressão Linear	
Valor Unitário	
Mínimo	499,16
Médio	533,88
Máximo	568,59
Valor Total	
Minimo (6,50%)	224.622,71
Médio	240.244,57
Máximo (6,50%)	255.866,42
Intervalo Predição	
Minimo (17,32%)	198.623,56
Máximo (17,32%)	281.865,57
Mínimo (IP)	441,39
Máximo (IP)	626,37
Campo de Arbítrio	
RL Mínimo	204.207,88
RL Máximo	276.281,25

Precisão = 13,0%

Regressão Linear	
Valor Unitário	
Mínimo	485,48
Médio	559,25
Máximo	638,23
Valor Total	
Minimo (13,19%)	218.466,89
Médio	251.662,43
Máximo (14,12%)	287.204,83
Intervalo Predição	
Minimo (18,99%)	203.861,58
Máximo (20,99%)	304.493,21
Mínimo (IP)	453,03
Máximo (IP)	676,65
Campo de Arbítrio	
RL Mínimo	213.913,07
RL Máximo	289.411,80

Precisão = 27,31%

Modelos com poucos dados MAIS SIMPLES são mais precisos do que os MAIS COMPLEXOS, entretanto, alguns pressuposto não são satisfeitos (como a heterocedasticidade).

Valor total =
 $-91746,98106$
 $+374,6959389 * \text{Area total}$
 $+307,6805557 * \text{Distancia ao polo valorizante}$

Valor total = (
 $-385,4159204$
 $+134,0835696 * \ln (\text{Area total})$
 $+0,0002409052179 * \text{Distancia ao polo valorizante}^2)$

CASO: INTERIOR DO ESTADO DO PARÁ

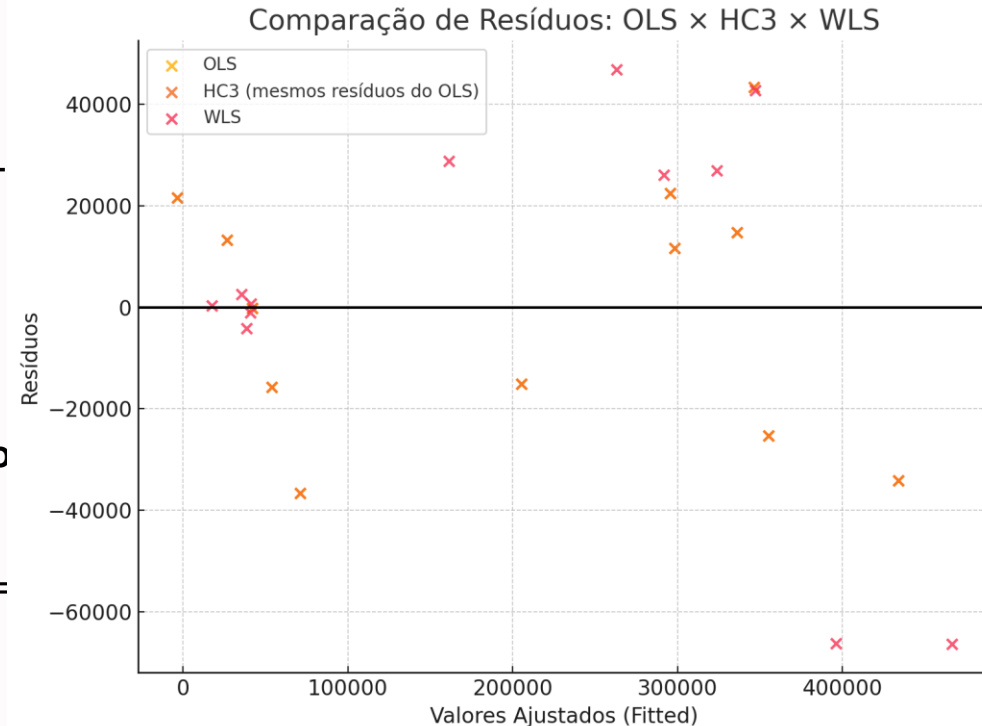
Erros-Padrão Robustos (HC3):-

Não altera coeficientes, apenas corrige as inferências;- A heterocedasticidade não invalida os coeficientes, mas distorce erros-padrão, testes t e IC. - Os erros robustos corrigem isso sem mexer no modelo.

Regressão Ponderada (WLS ou IRLS): - O método **WLS** corrige heterocedasticidade atribuindo **menor peso** às observações com maior variância dos erros. A variância dos resíduos aumenta com Y_i , então os pesos adotados foram $w_i = 1/Y_i^2$

$$Y^{\wedge} = -618,3390 + 580,4345X_1 + 3,5185X_2$$

Correção da heterocedasticidade tende a **melhorar a estabilidade**



CONSIDERAÇÕES FINAIS

•DESAFIOS PARA CONTORNAR PROBLEMAS

- Verificar se os pressupostos de Gauss-Markov foram atendidos (pelo que podem ser realizados ajustes, com transformação das variáveis)

- Enquadramento a NBR 14653 (Fundamentação e Precisão)

- Podem ser utilizadas técnicas mais robustas como

 - Bootstrap,

 - Validação cruzada,

 - Regressão Robusta – Método de Huber (RLM),

 - Regressão RANSAC (*Random Sample Consensus*),

 - Inferência Bayesiana,

- Especificamente para correção da heterocedasticidade

 - Erros-Padrão Robustos (HC3)

 - Regressão Ponderada (WLS ou IRLS)



AGRADECIMENTO



Engº. Civil **Érico Gaspar Lisboa**



erico.lisboa@unama.com.br



(91) 98237-6612