

Seleção de modelos de regressão e testes estatísticos nas avaliações pelo método comparativo

Ana Maria de Biazzi Dias de Oliveira

Arquiteta e Urbanista

O desafio de selecionar o modelo

mais adequado

A escolha não se limita apenas a identificar variáveis, mas equilibrar critérios como simplicidade, capacidade preditiva e robustez estatística

A IA ampliou exponencialmente a capacidade de coleta e processamento de dados, e transforma modelos estatísticos em laudos confiáveis, impõe avanço na aplicação de testes mais robustos no contexto das normas técnicas.

E este é o propósito da apresentação brilhante do eng. Antonio Pelli que minuciosamente nos traz à **com muita propriedade requisitos do Anexo A da ABNT NBR 14.653-2**, ainda não explorados em nossos laudos de avaliação.

A influencia da filosofia de Tukey

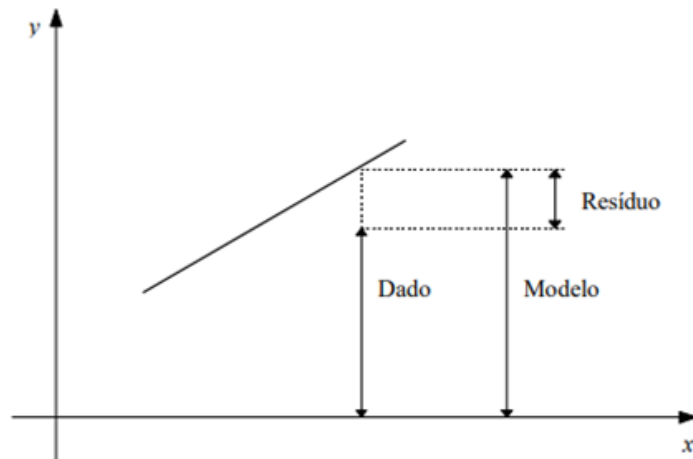
Fórmula base:

$$\text{Dado} = \text{Modelo} + \text{Resíduo}$$

Observações
reais
(valores
medidos)

Parte “suave”
dos dados -
estrutura
explicável,
padrão,

Parte
“grosseira” —
ruído, variação
não explicada,
exceções



Na regressão, por exemplo:

✓ **Modelo (M):** $\hat{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$

Representa o que conseguimos explicar com
variáveis, regressão..

✓ **Resíduo (R):** $\varepsilon = y - \hat{y}$

mostra o que **não conseguimos explicar** — e é onde
estão os **outliers, erros, heterocedasticidade, não-
linearidades.**

Tukey, John W et al. 1977. *Exploratory Data Analysis*. Vol. 2. Reading, MA.

Análise exploratória de dados

o ponto de partida

“Antes de ajustar qualquer modelo, precisamos entender os dados.”

- **Verificações iniciais:**

- Correlações entre variáveis explicativas (X).
- Distribuição da variável dependente (Y)
- Presença de outliers e multicolinearidade.

“Essa etapa é essencial para evitar modelos mal especificados, como alerta o teste RESET de Ramsey.”

“A análise exploratória também nos ajuda a decidir qual variável dependente usar: valor total ou valor unitário.”

Especificação da variável dependente

– Total ou Unitário?

Para a especificação correta da **variável dependente**, é necessária uma investigação no mercado em relação à **sua conduta e às formas de expressão dos preços** (por exemplo, **preço total ou unitário**). *ABNT NBR 14653-2- item 8.2.1.2)*



A norma **não impõe qual usar**, mas exige **coerência com o mercado e significância estatística dos parâmetros.**

A função é **inversa em relação à área**:
quanto maior o denominador, menor o resultado.
(**inversão funcional**)

Valor unitário (R\$/m²)

$$VU = \frac{VT}{A}$$

O valor total cresce proporcionalmente à área (função direta).
Mas o valor unitário é o quociente, logo a relação se torna inversa.

Exemplo 1: Avaliação de Terreno

1º opção Y - Valor Total

$$VT = -20.3851,71 + 2.582,80 * \text{Local} + 647,06 * \text{Área}$$

t=3,8 Sig 6%

t=61,65 sig 0,0001%

	x1	x2	Y	
	Local	Área Terreno	Valor Total	Valor unitário
	Proxy (I.F)	Quant. (m2)		
PROJEÇÕES	760	70.000	R\$ 46.148.573	R\$ 659
	210	70.000	R\$ 45.971.745	R\$ 657
	60	70.000	R\$ 45.360.884	R\$ 648
	210	15.000	R\$ 10.385.019	R\$ 692
	210	40.000	R\$ 26.560.804	R\$ 664

2º opção Y – Valor Unitário

$$V \text{ Unit} = +318,85 + 3,81 * \text{Local} + 146842,43 / \text{Área}$$

t=11,67 Sig 0,01%

t=2,11 Sig 4,27%

	x1	x2	Y	
	Local	Área Terreno	Valor Unitario	Valor Total
	Proxy (I.F)	Quant. (m2)		
PROJEÇÕES	760	70.000	R\$ 3.220	R\$ 225.423.800
	210	70.000	R\$ 1.122	R\$ 78.547.000
	60	70.000	R\$ 550	R\$ 38.489.500
	210	15.000	R\$ 1.130	R\$ 16.946.850
	210	40.000	R\$ 1.124	R\$ 44.946.800

- ✓ Área explica quase toda a variação do valor.
- ✓ R² elevado (>90%), indicando forte relação direta.
- ✓ t-Student de magnitude elevada, com p-valor muito abaixo de zero.
- ✓ Para uma mesma área, o valor projetado é **o mesmo em qualquer local**.

- ✓ A explicação compartilhada pela variação da área da localização.
- ✓ Valores projetados para mesma área variando em função da localização

Exemplo 2: Avaliação de Casas

1º opção Y - Valor Total

Valor total = $-665792.27 + 4390.54 * \text{Area} + 297520.79 * \text{Padrão}$
 $+ 117092.28 * \text{Suítes} + 143427.9679 * \text{Banheiros}$

Area	Padrão	Suítes	Banheiros	Valor unitário	Valor total
300	3	1	2	R\$ 9.856	R\$ 2.956.695
300	3	1	3	R\$ 9.042	R\$ 2.712.459
300	3	1	4	R\$ 8.227	R\$ 2.468.220
300	3	2	2	R\$ 9.303	R\$ 2.790.930
300	3	2	3	R\$ 8.489	R\$ 2.546.691
300	3	2	4	R\$ 7.675	R\$ 2.302.455
300	3	3	2	R\$ 8.751	R\$ 2.625.165
300	3	3	3	R\$ 7.936	R\$ 2.380.926
400	3	4	4	R\$ 6.401	R\$ 2.560.292

2º opção Y – Valor Unitário

Valor unitário = $+10180.79 - 2.317 * \text{Area} + 964.5502634 * \text{Padrão}$
 $- 681.25 * \text{Suítes} - 872.32 * \text{Banheiros}$

Valor total	Valor unitário
R\$ 1.853.776	R\$ 6.179
R\$ 2.004.785	R\$ 6.683
R\$ 2.155.794	R\$ 7.186
R\$ 1.925.560	R\$ 6.419
R\$ 2.076.569	R\$ 6.922
R\$ 2.227.578	R\$ 7.425
R\$ 1.967.550	R\$ 6.559
R\$ 2.118.559	R\$ 7.062
R\$ 2.757.575	R\$ 6.894

O valor total aumenta proporcionalmente à área (função direta).
A área na relação inversa traz suítes e banheiros e suítes também diminuindo valor unitário conduz a valores equivocados

Os pressupostos **básicos** modelo de regressão linear

garantir :

- ✓ Linearidade entre variáveis.
- ✓ **Normalidade dos resíduos.**
- ✓ **Homoscedasticidade, ou seja, variância constante dos erros.**
- ✓ Independência dos resíduos.
- ✓ Ausência de multicolinearidade entre preditores. Se esses pressupostos não forem atendidos, o modelo pode gerar conclusões equivocadas.



Quando esses pressupostos são violados, os intervalos de confiança e testes de significância podem ser distorcidos



são os primeiros alertas na validação do modelo

Normalidade e Homoscedasticidade

Verificação da Normalidade

Pelo exame do histograma , pela análise gráfica de resíduos padronizados versus valores ajustados ou da comparação da dos resíduos -1 ; $+1$ $-1,64$; $+1,64$ e $-1,96$; $+1,96$ nos intervalos de 68%, 90 % e 95 %;

A verificação da homoscedasticidade

pela análise gráfica dos resíduos versus valores ajustados,

ABNT NBR 14653-2

diagnósticos iniciais, mas que não oferecem soluções estruturais



Identificada violação do pressuposto

- ✓ Qual o efeito?
- ✓ Posso relaxar alguma hipótese?

diante de limitações interpretativas e resultados nem sempre conclusivos ou suficientes para validar o modelo

A apresentação introduziu muito bem os fundamentos da transformação Box-Cox, destacando sua função de estabilizar variância (homoscedasticidade) e aproximar normalidade

Ela ajusta a variável dependente por **meio de uma transformação de potência**, tornando o modelo mais consistente com os pressupostos da regressão linear

Box-Cox não apenas corrige, mas propõe uma nova escala de leitura da variável dependente

Transformação Box-Cox

Fórmula: $y' = \frac{y^{\lambda}-1}{\lambda}$ para $\lambda \neq 0$; se $\lambda = 0$, usa-se .

Estimação de λ : **máxima verossimilhança**.

•Casos particulares:

- $\lambda = -1$: inversa
- $\lambda = 0$: logarítmica
- $\lambda = 0,5$: raiz quadrada
- $\lambda = 1$: identidade (sem transformação)

•Cuidados:

exige $Y > 0$; se não, deslocar os dados.

Diagnóstico pós-transformação: ainda é necessário verificar normalidade e homoscedasticidade

Valores de λ na transformação Box Cox

Valor de λ	Transformação aplicada	Interpretação típica
-1	$1/Y$	Corrige forte assimetria positiva
-0,5	$1/\sqrt{Y}$	Reduz dispersão extrema
0	$\log(Y)$	Corrige crescimento exponencial
0,5	$\sqrt{Y}$	Suaviza contagens ou dados tipo Poisson
1	Y	Sem transformação (dados já bem comportados)

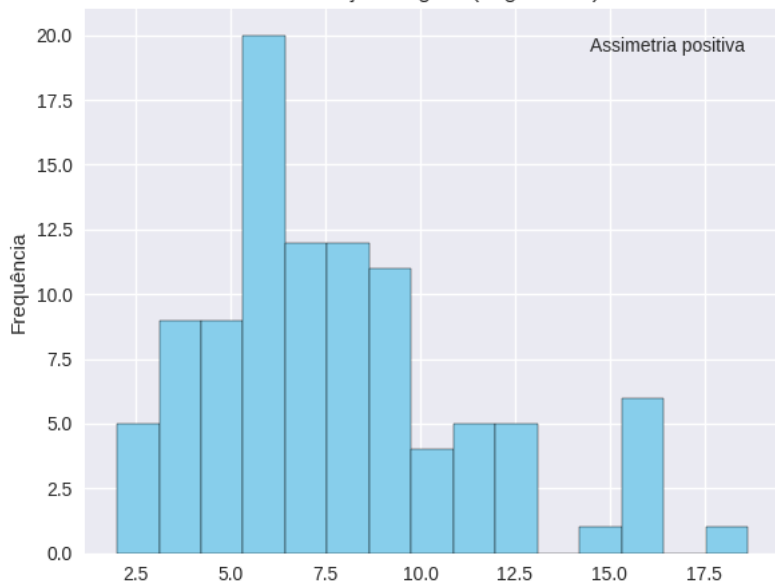
$\lambda < 1$: indica que os dados têm **assimetria positiva** ou variância crescente — a transformação suaviza essa distorção.

eficaz para estabilizar variância e normalizar distribuições com cauda longa.

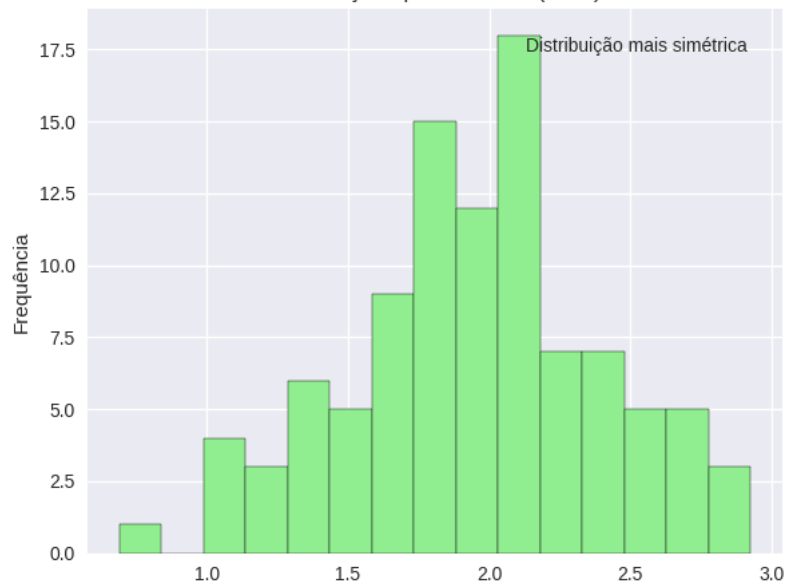
• $\lambda > 1$: útil quando há **assimetria negativa** ou compressão de valores altos.

- ✓ Para comparar modelos com e sem transformação, é importante usar o **AIC ajustado com Jacobiano**, garantindo que a comparação seja justa entre escalas diferentes.
- ✓ A escolha final do modelo deve equilibrar **qualidade estatística (AIC/BIC, R^2 ajustado)** e **interpretação prática** dos coeficientes.

Distribuição Original (Log-normal)



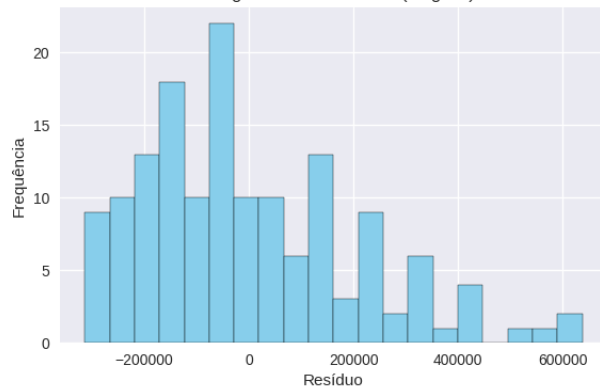
Distribuição após Box-Cox ($\lambda = 0$)



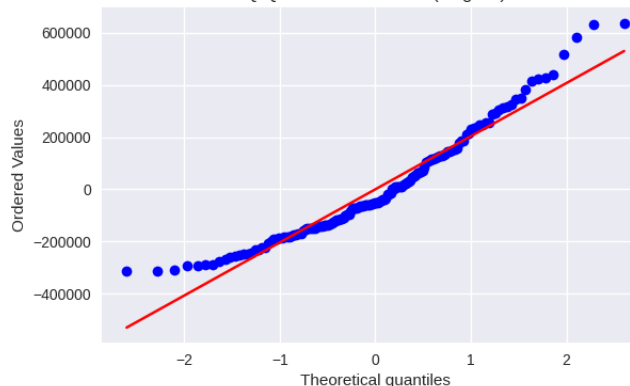
- **Antes da transformação:** distribuição com forte assimetria positiva.
- **Depois da Box-Cox ($\lambda = 0$):** distribuição mais simétrica e com variância estabilizada.

Transformação Box-Cox

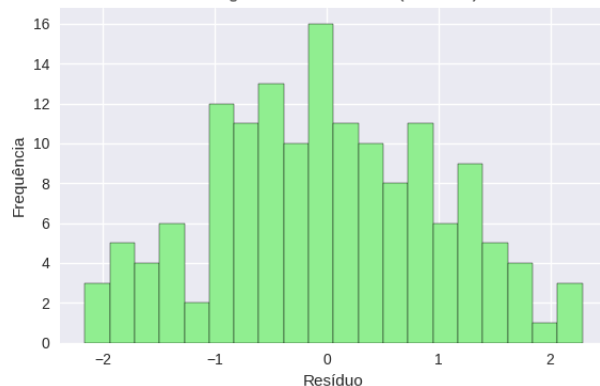
Histograma dos Resíduos (Original)



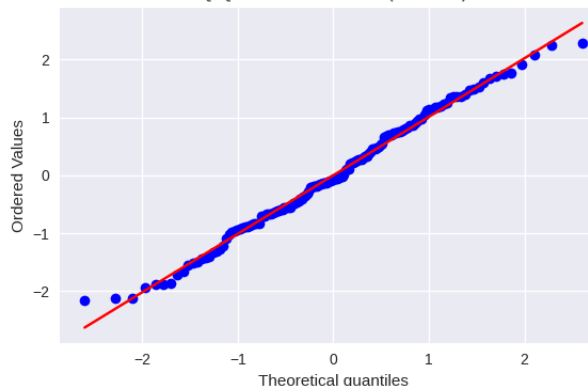
Q-Q Plot dos Resíduos (Original)



Histograma dos Resíduos (Box-Cox)



Q-Q Plot dos Resíduos (Box-Cox)



A transformação Box-Cox ajuda a **normalizar a variável resposta** e estabilizar a variância, melhorando a confiabilidade das inferências

O que muda com a IA nas análises de regressão linear?



Ana Maria de Biazzi Dias de Oliveira
Arquiteta de Urbanista



(55) 11 98109-7733



anabiazzi@uol.com.br