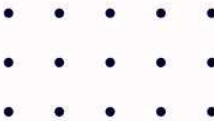


SELEÇÃO DE MODELOS DE REGRESSÃO E TESTES ESTATÍSTICOS NAS AVALIAÇÕES PELO MÉTODO COMPARATIVO

Lutemberg de Araújo Florencio, Ph.D.

□ Estrutura da apresentação

1. Motivação e contextualização
2. Fundamentos (modelos)
3. Pressuposto da linearidade



1. Motivação

✓ Tudo isto parece ser **muito complexo**.

De fato, a Engenharia de Avaliações é complexa e por isso requer profissionais especialistas.



Roteiro

Introdução à Regressão Linear – NBR 144.653-2

Problema da seleção de variáveis

Métodos: Forward, Backward, Stepwise

Crítérios: R^2 , R^2 ajustado, AIC, BIC, Cp

Teste F e ANOVA

Pressupostos e testes

Diagnóstico de resíduos e influência

Multicolinearidade (VIF)

Validação cruzada

Recomendações práticas

Seleção de modelos de regressão linear múltipla e testes estatísticos sugeridos

Introdução à Regressão Linear Múltipla

A regressão linear múltipla é uma técnica estatística que permite analisar a relação entre uma variável dependente (Y) e múltiplas variáveis independentes (X).

$$Y = X\beta + \epsilon, \text{ com } E(\epsilon)=0, \text{ Var}(\epsilon)=\sigma^2 I$$

Onde:

Y é o vetor da variável resposta

X é a matriz de variáveis preditoras

β é o vetor de parâmetros a estimar

ϵ são os erros aleatórios

Anexo A

A.3 Testes de Significância

A.3.1 O nível de significância máximo admitido nos demais testes estatísticos (aqueles não citados na Tabela 1) não deve ser superior 10%.

A.3.2 A significância de subconjuntos de parâmetros, quando pertinente, pode ser testada pela análise da variância por partes.

A.3.3 Os níveis de significância utilizados nos testes citados em A.3 serão compatíveis com a especificação da avaliação.

A qualidade do modelo depende da seleção adequada de variáveis e da verificação dos pressupostos estatísticos.

Recomendações Práticas

Fluxo de trabalho sugerido para análise de regressão:

- 1 Exploração e pré-processamento: análise descritiva, correlações, transformações e tratamento de dados
- 2 Especificação inicial: identificar possíveis termos e interações com base em teoria e visualização
- 3 Ajuste e diagnóstico completo: verificar linearidade, normalidade, homocedasticidade, independência
- 4 Tratamento de violações: transformações, erros-padrão robustos, modelos alternativos (GLM, etc.)
- 5 Seleção de variáveis: critérios objetivos (AIC/BIC/penalização) + validação cruzada
- 6 Verificação de estabilidade: bootstrap para coeficientes, análise de multicolinearidade
- 7 Relatório final: coeficientes, intervalos de confiança, métricas de desempenho e limitações

Observação: manter variáveis de controle teoricamente importantes mesmo quando não significativas estatisticamente. O conhecimento deve complementar a análise quantitativa.

Seleção de modelos de regressão linear múltipla e testes estatísticos sugeridos

AIC, BIC e Cp de Mallows

Crítérios de informação: Balanceiam a qualidade do ajuste com o número de parâmetros.

AIC (Akaike Information Criterion):

$$AIC = -2\log(L) + 2p$$

$$AIC = n \log(SQRes/n) + 2p$$

BIC (Bayesian Information Criterion):

$$BIC = -2\log(L) + p \log(n)$$

$$BIC = n \log(SQRes/n) + p \log(n)$$

Cp de Mallows:

$$Cp = SQRes/\sigma^2 - (n - 2p)$$

Objetivo:

Minimizar AIC e BIC

Cp próximo a p indica modelo não enviesado

Comparação entre critérios:

Critério	Penalização	Preferência	Amostra
AIC	2p	Moderada	Qualquer tamanho
BIC	$p \log(n)$	Alta	Grandes ($n > 100$)
Cp	2p	Moderada	Qualquer tamanho
R^2 Ajustado	$(n-1)/(n-p)$	Baixa	Qualquer tamanho

Principais diferenças:

BIC penaliza mais modelos complexos

BIC favorece modelos mais simples que AIC

Cp considera explicitamente o viés do modelo

Consistência assintótica: BIC > AIC

Pressupostos do Modelo e Consequências

1. Linearidade

Relação linear entre preditores e variável resposta

Consequência: Previsões enviesadas, subestimação ou superestimação

Correção: Transformações (log, Box-Cox).

2. Normalidade dos Resíduos

Erros aleatórios com distribuição normal

Consequência: Testes t e F inválidos; ICs imprecisos

Correção: Transformação de variáveis; bootstrap

3. Homocedasticidade

Variância constante dos erros

Consequência: Estimadores inefficientes; testes estatísticos inválidos

Correção: erro-padrão robusto (Park, White)

4. Independência

Ausência de autocorrelação nos resíduos

Consequência: Subestimação dos erros-padrão; inflação do R^2

Correção: erro-padrão robusto

5. Ausência de Multicolinearidade

Preditores não altamente correlacionados

Consequência: Coeficientes instáveis; erros-padrão inflacionados

Correção: remover/combina variáveis

6. Ausência de Pontos Influentes

Observações não devem ter influência desproporcional

Consequência: Distorção nos coeficientes; resultados não generalizáveis

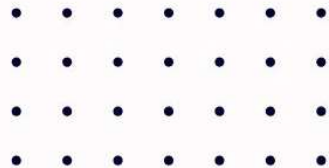
Correção: Regressão robusta; análise de influência; validação cruzada

1. Motivação



Eng.º Domingos de Saboya

*Quem avalia necessita ter fortemente desenvolvido o **sentido da pesquisa** para levar a cabo tal serviço com seriedade. Se for intolerante à **prática da pesquisa** não poderá avaliar, ou pelo menos será melhor que não o faça.*
(Extraído dos Classificados do JORNAL DO BRASIL, de 30.AGO.87)



SELEÇÃO DE **MODELOS** DE REGRESSÃO E TESTES ESTATÍSTICOS NAS AVALIAÇÕES PELO MÉTODO COMPARATIVO



❑ Fundamentos

✓ **Modelo:** representação simplificada da realidade.

Desenho



❑ Fundamentos

Método comparativo direto de dados de mercado

✓ **Modelo:** representação simplificada do comportamento da realidade do mercado imobiliário (analisado/estudado), **expressa a partir de uma equação matemática.**

PERGUNTA: Como seria esta “equação matemática”?



1º PROBLEMA

❑ Fundamentos

Método comparativo direto de dados de mercado

✓ **Teoria dos preços hedônicos** [Lancaster & Rosen (1974)]:

O preço do bem pode ser decomposto nos preços de suas k **características**, representadas por $q = (q_1, q_2, \dots, q_k)$, por meio de uma função do tipo:

$$P(q) = P(q_1, q_2, \dots, q_k).$$

Locacionais, físicas e econômicas

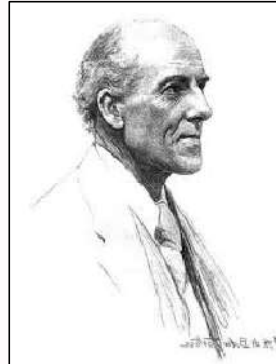
DESAFIO:

(I) **Quais** variáveis devem ser consideradas?



2º PROBLEMA

❑ Fundamentos



Karl Pearson

Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia

Karl Pearson

Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character

Vol. 187 (1896), pp. 253-318 (66 pages)

✓ **Pearson** desenvolveu e formalizou o conceito de correlação como conhecemos hoje.

IMPORTANTE: Pearson mede apenas correlação linear e **não implica relação de CAUSALIDADE entre as variáveis.**

❑ Fundamentos

Estudo revela que decorar jardim com gnomos e duendes desvaloriza imóvel vizinho

Fonte: <https://casavogue.globo.com/>



Na Europa, uma pesquisa revelou que esses enfeites desvalorizam a casa do vizinho em até £12 mil (cerca de R\$ 90 mil, na cotação atual).

❑ Fundamentos

Método comparativo direto de dados de mercado

✓ **Teoria dos preços hedônicos** [Lancaster & Rosen (1974)]:

O preço do bem pode ser decomposto nos preços de suas k **características**, representadas por $q = (q_1, q_2, \dots, q_k)$, por meio de uma função do tipo:

$$P(q) = P(q_1, q_2, \dots, q_k).$$

Locacionais, físicas e econômicas

DESAFIO:

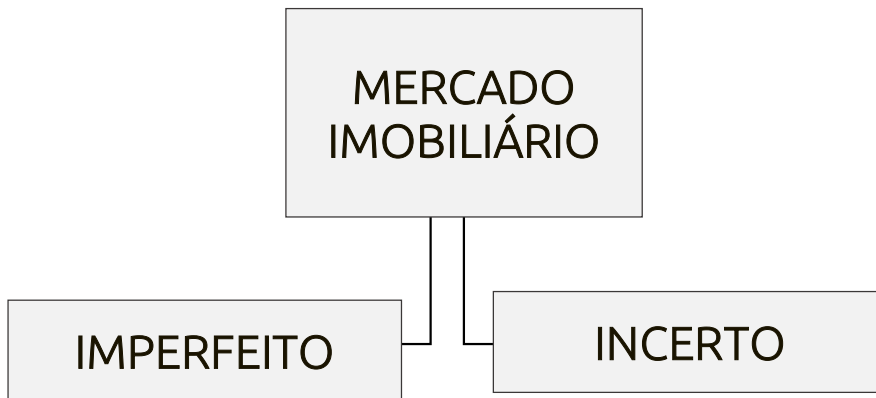
(I) **Como** obter esta equação (**modelo**)?



3º (BOM)
PROBLEMA

❑ Fundamentos

Método comparativo direto de dados de mercado

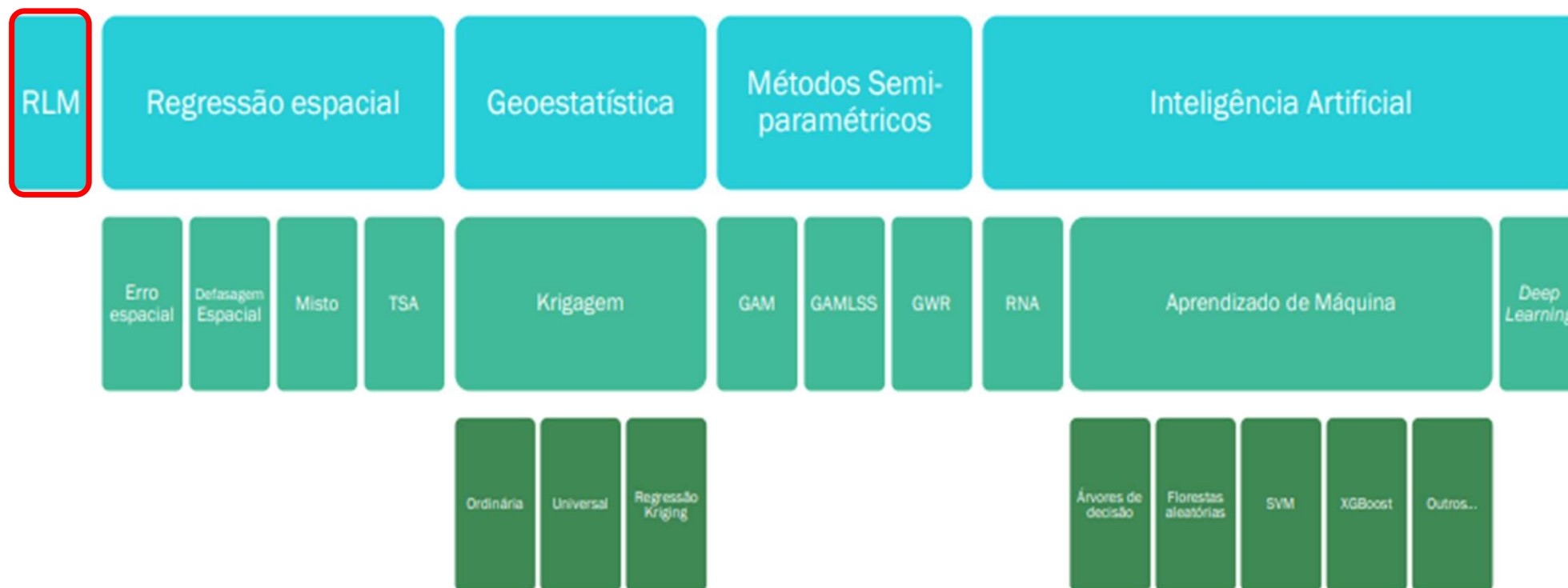


✓ **Ou seja:** o avaliador geralmente se depara com uma amostra constituída por elementos pesquisados com **características heterogêneas** entre si e em relação ao bem avaliando.

CONCLUSÃO: O uso da média aritmética é **INADEQUADO**.

❑ Fundamentos

AVALIAÇÃO DE IMÓVEIS – Método comparativo



PERGUNTA: COMO obter a equação (modelo)?
Depende da **ferramenta** (técnica) estatística escolhida.

Fonte: MSc. Augusto Oliveira (apresentação R)

❑ Fundamentos



Anexo A
(normativo)

Procedimentos para a utilização de modelos de regressão linear

✓ A preocupação da ABNT NBR 14653 2/3 não está apenas na obtenção do modelo **mas também em sua VALIDAÇÃO**.

✓ De acordo com a NBR 14653-2/3, no tratamento dos dados podem ser utilizados, alternativamente e em função da qualidade e da quantidade de dados e informações disponíveis:

- **Tratamento científico:** tratamento de evidências empíricas pelo uso de metodologia científica que leve à indução de **modelo validado** para o comportamento do mercado.

PERGUNTA: O que vem a ser um **modelo validado**?

❏ Fundamentos

1. O **MODELO** deve:

(i) apresentar adequação teórica e lógica com o segmento de mercado analisado/estudado.

Modelo: representação simplificada do comportamento da realidade do mercado imobiliário (analisado/estudado), **expressa a partir de uma equação matemática.**

❑ Fundamentos

2. (ii) O **MODELO** deve atender aos pressupostos/premissas/requisitos exigidos pela CLASSE DE MODELOS utilizada.

Pressupostos do Modelo e Consequências	
1. Linearidade Relação linear entre preditores e variável resposta <i>Consequência: Previsões enviesadas, subestimação ou superestimação</i> Correção: Transformações (log, Box-Cox).	2. Normalidade dos Resíduos Erros aleatórios com distribuição normal <i>Consequência: Testes t e F inválidos; ICs imprecisos</i> Correção: Transformação de variáveis; bootstrap
3. Homocedasticidade Variância constante dos erros <i>Consequência: Estimadores ineficientes; testes estatísticos inexatos</i> Correção: erro-padrão robusto (Park, White)	4. Independência Ausência de autocorrelação nos resíduos <i>Consequência: Subestimação dos erros-padrão; inflação do R^2</i> Correção: erro-padrão robusto
5. Ausência de Multicolinearidade Preditores não altamente correlacionados <i>Consequência: Coeficientes instáveis; erros-padrão inflacionados</i> Correção: remover/combinar variáveis	6. Ausência de Pontos Influentes Observações não devem ter influência desproporcional <i>Consequência: Distorção nos coeficientes; resultados não generalizáveis</i> Correção: Regressão robusta; análise de influência; validação cruzada

Fonte: PELLÍ NETO, A. (XXIII XOBREAP)

❑ Pressuposto da linearidade



✓ O pressuposto da linearidade afirma que a relação entre a variável dependente (resposta) e as variáveis independentes (explicativas) pode ser adequadamente representada por uma **combinação linear dos parâmetros** do modelo de regressão.

✓ O nome “regressão linear” **não exige** que a relação entre X e Y seja uma “reta” nos dados originais, mas que seja uma **função linear no(s) parâmetro(s)** e que a **relação média** entre elas pode ser bem modelada por uma reta (ou hiperplano, no caso múltiplo).

- $Y = \beta_0 + \beta_1 X + \varepsilon$
- $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$
- $Y = \beta_0 + \beta_1 \log(X) + \varepsilon$

❑ Pressuposto da linearidade

A.2.1 Verificação dos pressupostos do modelo

A.2.1.1 Linearidade

Recomenda-se que seja analisado primeiramente o comportamento gráfico da variável dependente em relação a cada variável independente, em escala original. Isto pode orientar o avaliador na transformação a adotar. Existem formas estatísticas de se buscar a transformação mais adequada, como, por exemplo, os procedimentos de Box e Cox.

As transformações utilizadas para linearizar o modelo devem, tanto quanto possível, refletir o comportamento do mercado, com preferência pelas transformações mais simples de variáveis, que resultem em modelo satisfatório.

Após as transformações realizadas, se houver, examina-se a linearidade do modelo, pela construção de gráficos dos valores observados para a variável dependente *versus* cada variável independente, com as respectivas transformações.

❑ Pressuposto da linearidade



✓ Métodos para verificação do pressuposto da linearidade:

1. Análise gráfica (resíduos x valores ajustados x variáveis independentes):
ausência de padrões sistemáticos
2. Teste RESET – Ramsey (*): teste de hipótese
3. Análise gráfica dos **resíduos parciais**

❑ Pressuposto da linearidade

✓ 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:

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

- Para obter resíduos parciais em relação a uma variável explicativa, retornamos o efeito dela à equação acima:

- $\hat{\varepsilon}_1 = Y - \hat{Y} + \hat{\beta}_1 X_1$
- $\hat{\varepsilon}_2 = Y - \hat{Y} + \hat{\beta}_2 X_2$
- ...
- $\hat{\varepsilon}_k = Y - \hat{Y} + \hat{\beta}_k X_k$

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

❑ Pressuposto da linearidade

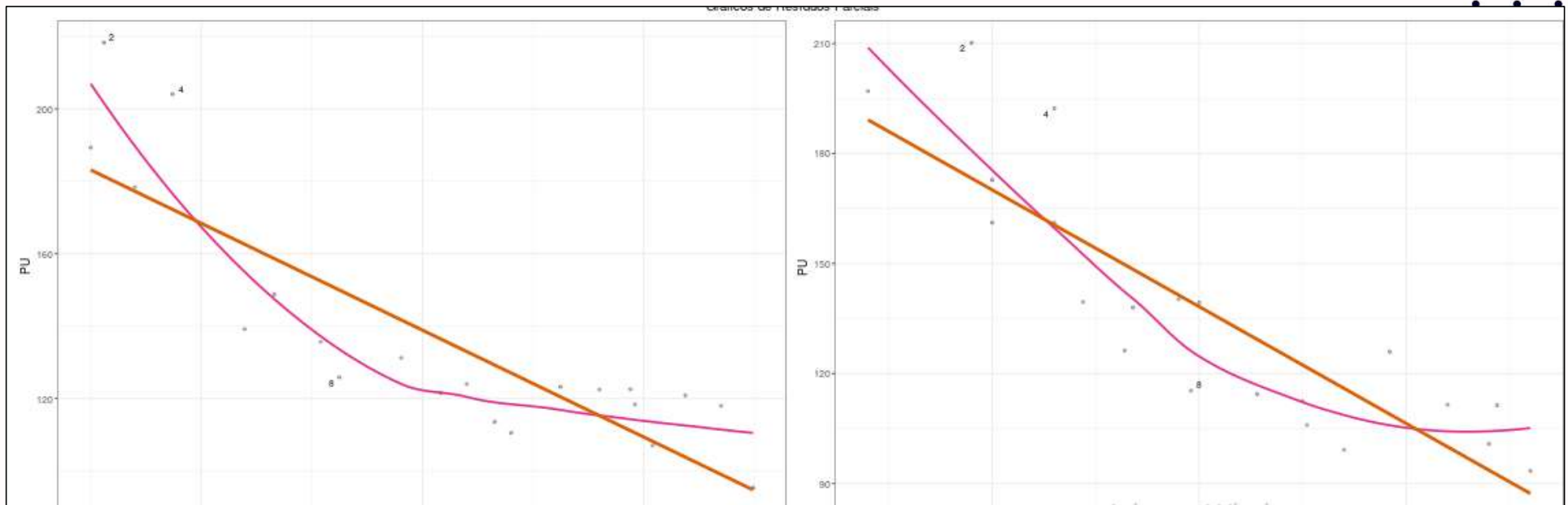
Dados

- 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

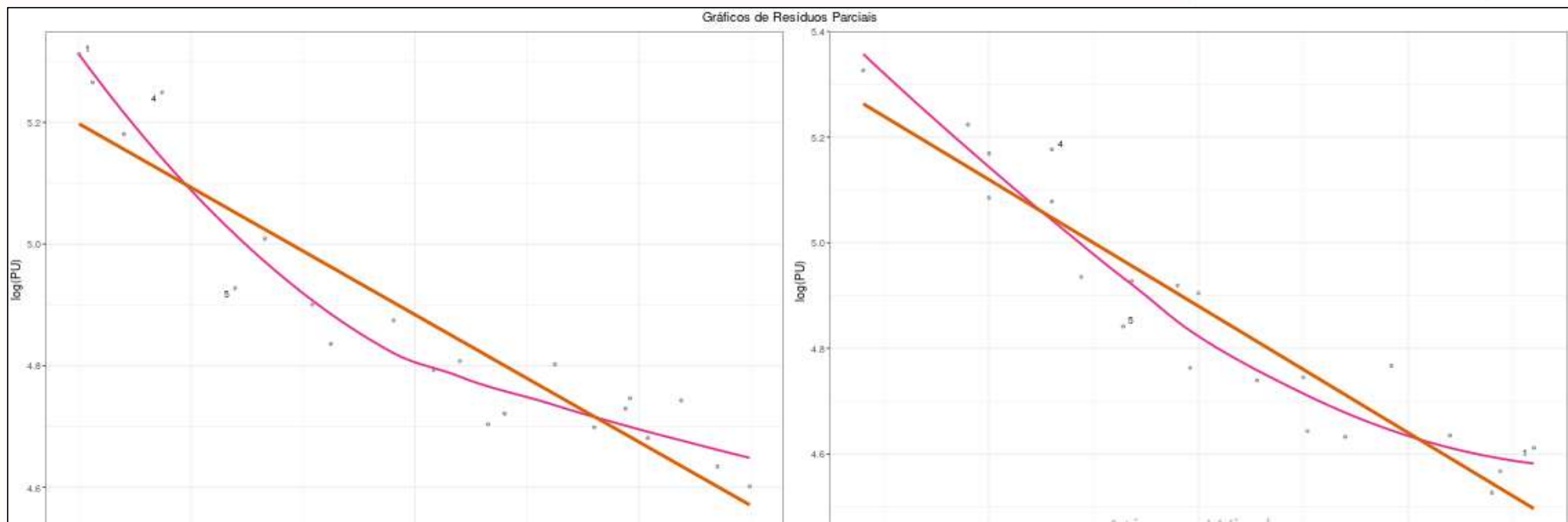
❑ Pressuposto da linearidade



$$PU \sim A+D$$

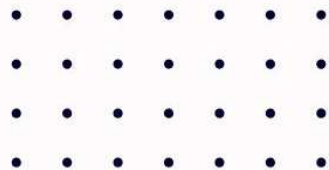
- Ajuste do modelo
- Regressão local

❑ Pressuposto da linearidade

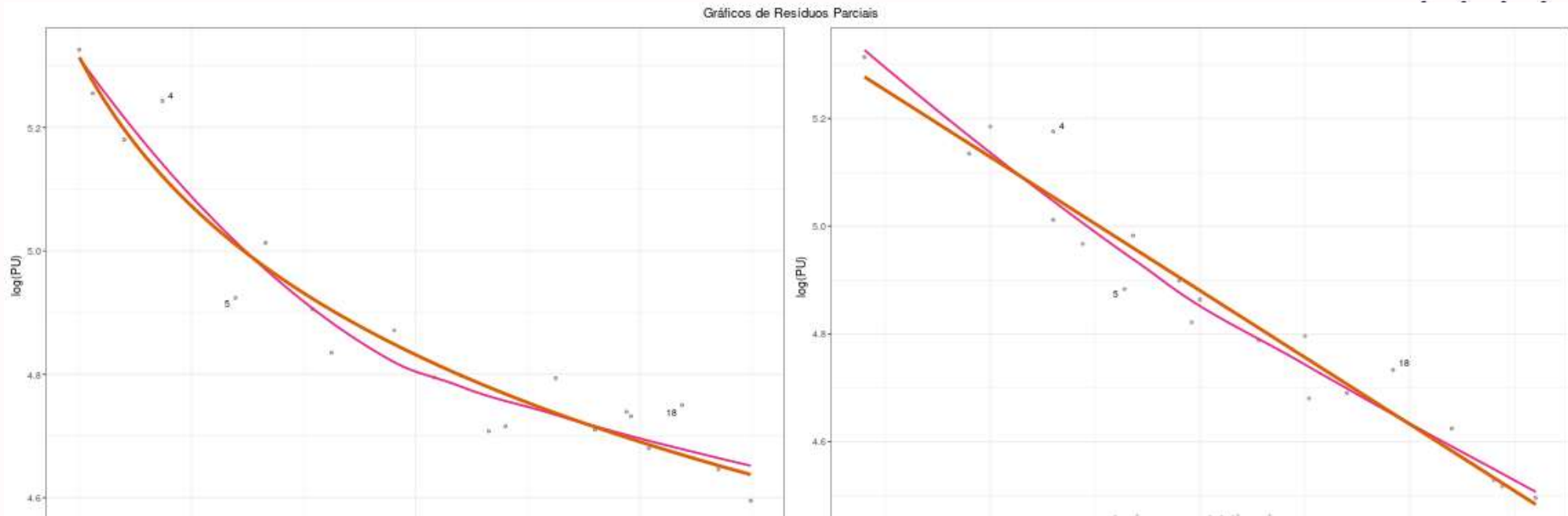


$$\text{Ln(PU)} \sim \text{A+D}$$

- Ajuste do modelo
- Regressão local

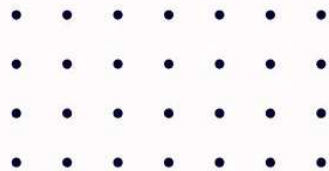


❑ Pressuposto da linearidade



$$\text{Ln}(\text{PU}) \sim \text{Ln}(\text{A}) + \text{D}$$

- Ajuste do modelo
- Regressão local



Obrigado!



@prof.lutembergflorencio



(81) 98886-1406



lutembergflorencio@yahoo.com.br