

# Particularidades da regressão linear



## Lutemberg Florencio

Doutor em Eng. Civil (*Real Estate*) – USP

Engenheiro Civil (POLI-UPE)

Espec. em Aval. e Perícias de Eng. (FOC-SP)

Mestre em Estatística (UFPE)

Eng. Civil do Banco do Nordeste (BNB)

Prof. da disc. de Eng. de Aval. da UNINASSAU

Vice-presidente técnico da SOBREA

Promoção:



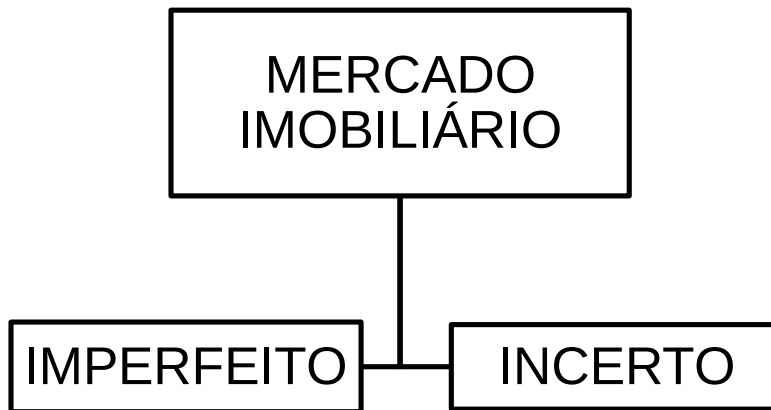
Realização:



# ❖ Motivação



- Orientando de R. Fisher
- Iniciou o aprendizado do “R” aos 95 anos.

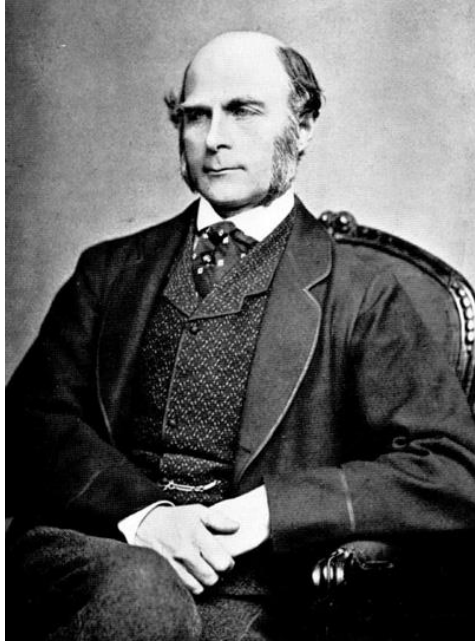


# ❖ Motivação



# ❖ Motivação

## ✓ Análise de regressão



Modelagem em que se busca compreender como o comportamento de uma(s) variável(is) [variáveis independentes] pode mudar o comportamento de uma outra variável [variável dependente].



# ❖ Motivação

## ✓ Teoria dos preços hedônicos (Lancaster & Rosen):

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



# ❖ Motivação

VD x VI

Análise de regressão

Modelo clássico de regressão linear (MCRL)

VD é expressa por uma **combinação linear nos parâmetros**, sendo dada por:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki} + \hat{e}_i$$

> Com base em uma amostra extraída do mercado, os parâmetros populacionais são estimados por inferência estatística (**MQO**).



# ❖ Contextualização

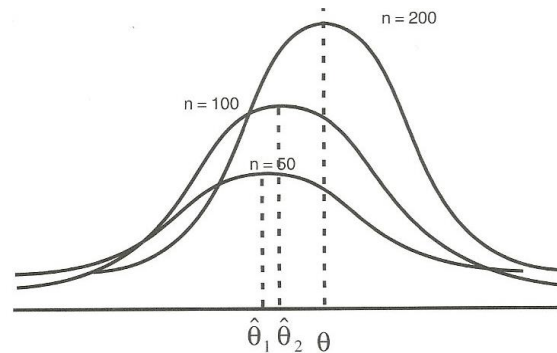
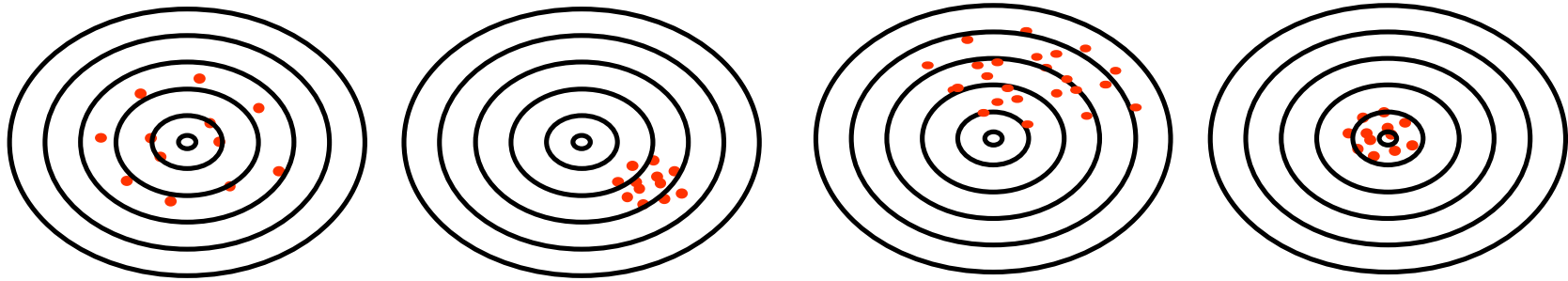
✓ Dada as premissas do modelo clássico de regressão linear, os estimadores MQO são BLUE -- Teorema de Gauss-Markov --.

- (i) Linearidade
- (ii) Homocedasticidade
- (iii) Não-multicolinearidade
- (iv) Não-autocorrelação
- (v) **Especificação correta do modelo**



# ❖ Contextualização

- ✓ Estimadores MQO: não-tendenciosidade, eficiência e consistência.



# ❖ Contextualização

✓ NBR 14653-2

**Anexo A**, item A.2.1 – Verificação dos pressupostos do modelo:

- (i) Linearidade
- (ii) Normalidade
- (iii) Homocedasticidade
- (iv) Não-multicolinearidade
- (v) Não-autocorrelação
- ~~(vi) Inexistência de pontos atípicos~~

Permite deduzir as distribuições de probabilidade dos parâmetros  $\hat{\beta}_i$  (normal) e de  $\sigma^2$  (qui-quadrado), fundamental para estabelecer intervalo de confiança e promover testes de hipóteses. Estimador MQO é BUE (Teorema de Rao).



# ❖ Contextualização

✓ NBR 14653-2

**Anexo A**, item A.2.1 – Verificação dos pressupostos do modelo:

- (i) Linearidade
- (ii) Normalidade
- (iii) Homocedasticidade
- (iv) Não-multicolinearidade
- (v) Não-autocorrelação
- ~~(vi) Inexistência de pontos atípicos~~

E o pressuposto referente à  
**correta especificação do  
modelo?**



# ❖ Correta especificação do modelo

✓ Ao formular um modelo é possível cometer um ou mais dos seguintes erros de especificação:

1. Omissão de variáveis
2. Inclusão de variáveis desnecessárias
3. Adoção de forma funcional errada
4. Erros de medição (regressando ou regressores)
5. Especificação incorreta do termo de erro estocástico



# ❖ Correta especificação do modelo

✓ Questões relevantes:

1. Quais as técnicas para seleção do modelo “correto”?
2. Quais as consequências de cada tipo de erro de especificação?
3. Como detectar erros de especificação?
4. Depois de detectar erros de especificação, que soluções podem ser adotadas?
5. Como avaliar o desempenho de modelos alternativos?

→ O que diz a NBR 14653-2?



❖ Correta especificação do modelo

- ✓ Detecção de erros de especificação:

## 1. Análise gráfica

> Gráfico dos resíduos x variável independente (padrões caract.)

## 2. Testes formais

### > Teste RESET (*Regression Specification Error Test*) de Ramsey

J.B. Ramsey (1969), Tests for Specification Error in Classical Linear Least Squares Regression Analysis. *Journal of the Royal Statistical Society*, Series B 31, 350–371



# ❖ Teste RESET

✓ Suponha que tenha sido especificado o modelo:

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + e_t$$

Sejam os valores preditos de  $y_t$ :

$$\hat{y}_t = b_1 + b_2 x_{t2} + b_3 x_{t3}$$

Consideremos os dois seguintes modelos “artificiais”:

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \gamma_1 \hat{y}_t^2 + e_t$$

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \gamma_1 \hat{y}_t^2 + \gamma_2 \hat{y}_t^3 + e_t$$



# ❖ Teste RESET

✓ Segundo Ramsey (1969), ao incluirmos  $\hat{y}_t$  como regressor (modelo artificial), tem-se um automático incremento do coeficiente de determinação ( $R^2$ ). Se o aumento de  $R^2$  for estatisticamente significativo (segundo o teste F), sugere-se que o modelo original esteja mal especificado.

$H_0$  = modelo sem erro de especificação;

\* Regra de decisão

$H_1$  = modelo com erro de especificação;

Se  $F_{calc} > F_{tab}$ , rejeita-se  $H_0$ .

$$F_{(k;n-k)} = \frac{(R_{novo}^2 - R_{velho}^2) / \text{número de novos regressores}}{(1 - R_{novo}^2) / (n - \text{número de parâmetros no novo modelo})}$$



# ❖ Aplicação



Detalhes sobre o projeto R, colaboradores, documentação e diversas outras informações podem ser encontradas na página oficial do projeto em: <http://www.r-project.org>.

XV COBREAP (2009)

*UTILIZAÇÃO DO SOFTWARE (GRATUITO) R  
NA ENGENHARIA DE AVALIAÇÕES*

Florencio, L. A.

### PACOTES UTILIZADOS ###

Library(lmtest) #teste RESET

##### Teste de RESET ###

```
resettest(formula, power = 2:3, type = c("fitted", "regressor", "princomp"), data = list())
```



# ❖ Aplicação

✓ Modelo para avaliação de uma casa residencial, sendo:

➤ Variável dependente: preço unitário

➤ Variáveis independentes: área construída e localização

✓ Modelo (1)

$$PU = \beta_0 + \beta_1 AC + \epsilon$$

## Teste RESET

```
resettest(PU ~ AC, power=2:3, type="fitted",  
data=dados)
```

P-valor: 0,029 --- Rejeita-se  $H_0$  ao nível de 5%, ou seja, o modelo apresenta erro de especificação.

✓ Modelo (2)

$$PU = \beta_0 + \beta_1 AC + \beta_2 LOC + \epsilon$$

## Teste RESET

```
resettest(PU ~ AC+LOC, power=2:3,  
type="fitted", data=dados)
```

P-valor: 0,78 --- Não há evidências para rejeição de  $H_0$ , ou seja, o modelo está corretamente especificado.



# ❖ Conclusão

| METODOLOGIA CIENTÍFICA                                | ENGENHARIA DE AVALIAÇÕES (MCDDM) |
|-------------------------------------------------------|----------------------------------|
| Observação do fenômeno                                | Vistoria                         |
| Preparação da pesquisa                                | Planejamento da pesquisa         |
| Condução de experimentos                              | Coleta de dados de mercado       |
| Processamento e análise dos dados                     | Análise exploratória e modelagem |
| Verificação do ajuste da técnica de análise escolhida | Validação do modelo.             |
| Redação do relatório de pesquisa.                     | Elaboração do Laudo de avaliação |





*A Estatística é a gramática da Ciência.*



# Obrigado!

✉ lutembergflorencio@yahoo.com.br

📷 @prof.lutembergflorencio

📞 (81) 98886-1406

