

MÓDULO III

# Regressão polinomial

**Alexandre Loureiros Rodrigues**

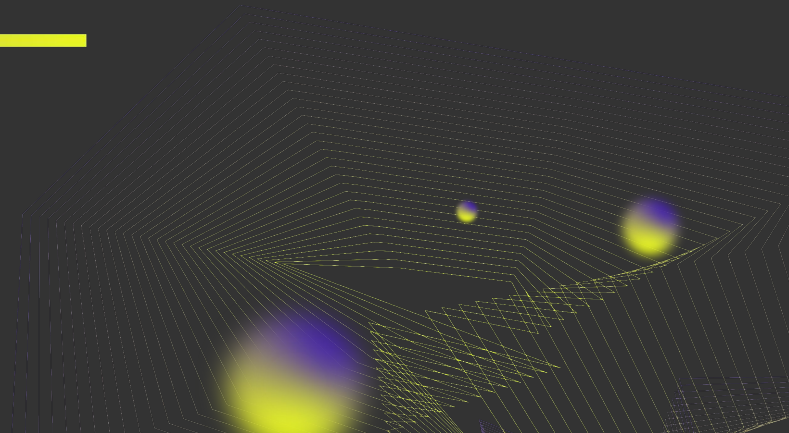
Professor do departamento de Estatística - UFES

ESPECIALIZAÇÃO

INTELIGÊNCIA ARTIFICIAL  
& CIÊNCIA DE DADOS

SEAD  
UFES

Superintendência de  
Educação a Distância



# ÍNDICE

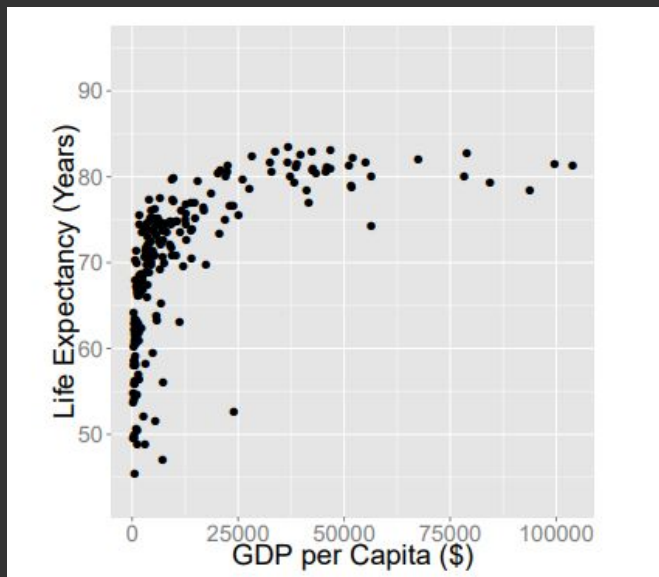


1. Introdução
2. Modelo polinomial
3. Seleção de modelos

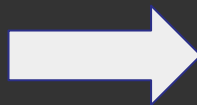


# 1. Introdução

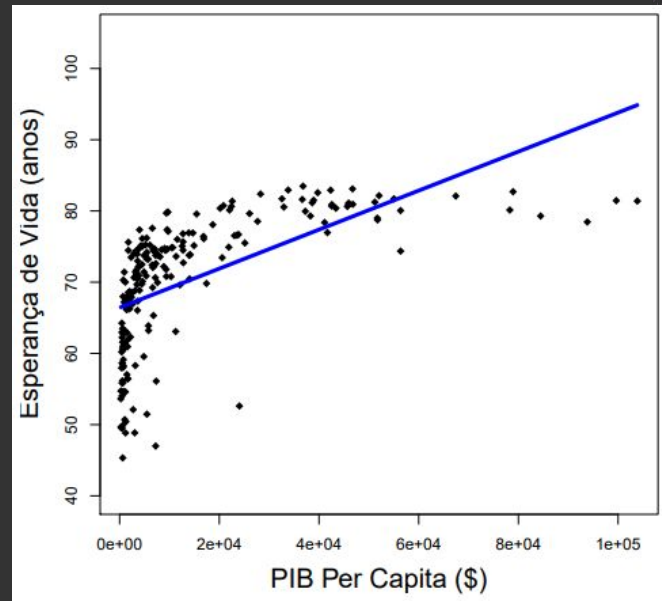
# Introdução - Exemplo



Ajuste linear



Péssimo  
poder  
preditivo





## 2. Modelo linear polinomial

# Formulação matemática

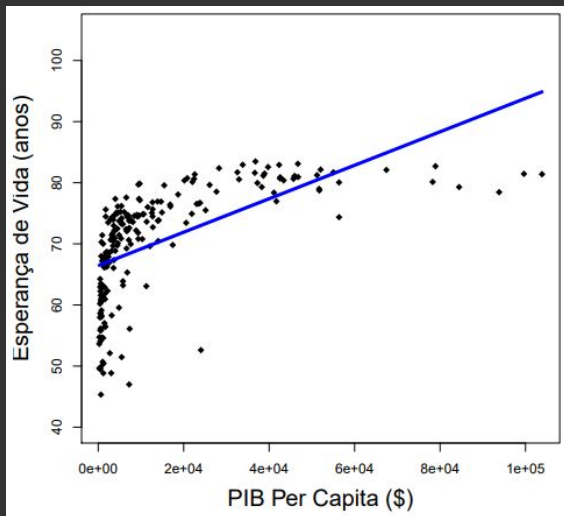
- Para problema mais complexos podemos ajustar polinômio

$$\hat{Y} = \theta_0 + \theta_1 X + \theta_2 X^2 + \dots + \theta_p X^p$$

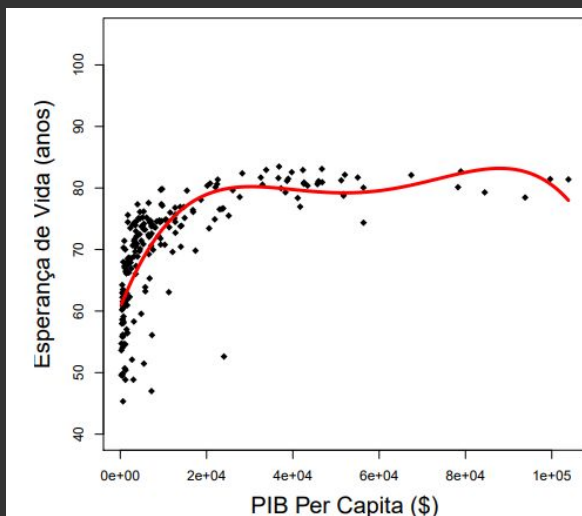
- **Ponto de atenção:** Polinômios de grau muito alto podem ajustar-se bem aos dados de treino, mas não generalizar para novos dados
- Estimação pode ser feita de forma similar ao caso simples  
Encontrar  $\theta_0, \theta_1, \dots, \theta_p$  que minimize

$$\sum_{i=1}^n L_i = \sum_{i=1}^n \left( \hat{Y}_i - Y_i \right)^2$$

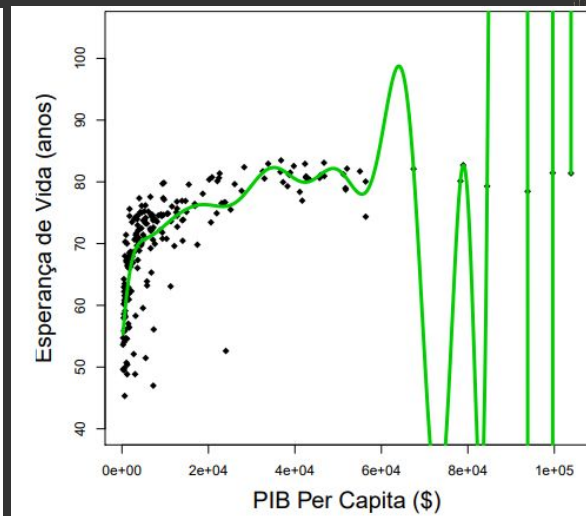
# Voltando ao exemplo



$$p = 1$$



$$p = 4$$

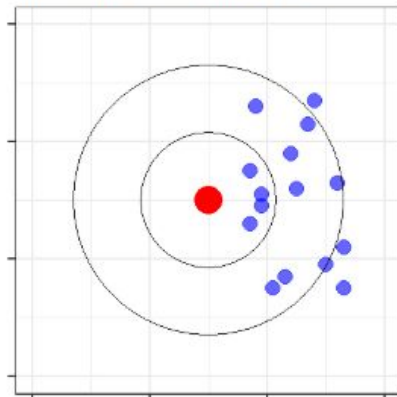


$$p = 50$$

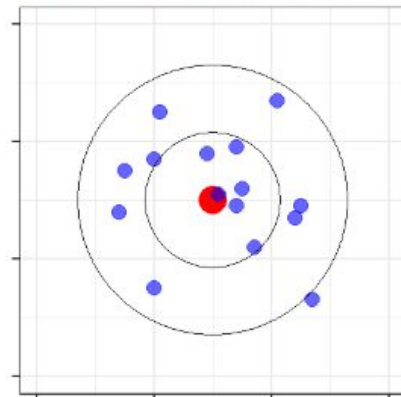
Como comparar os modelos e achar o melhor  $p$  ?

# Vício X Variância

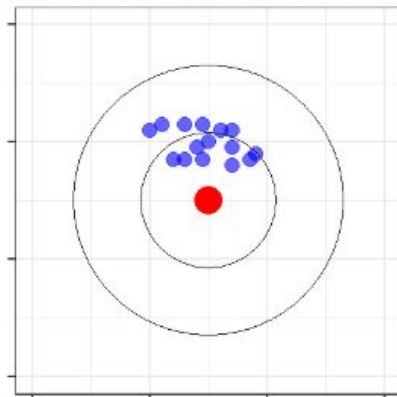
(a) Vício grande e Variância grande



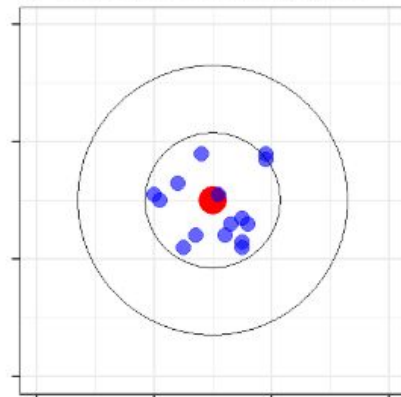
(b) Vício pequeno e Variância grande



(c) Vício grande e Variância pequena

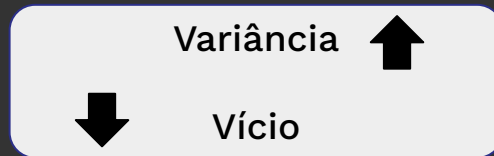
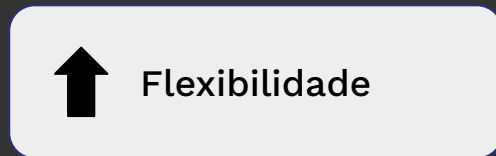
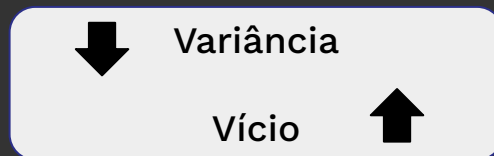
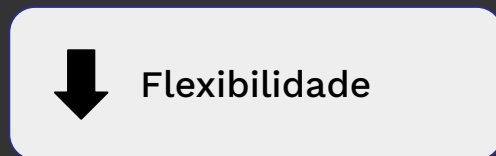


(d) Vício pequeno e Variância pequena





# Vício X Variância

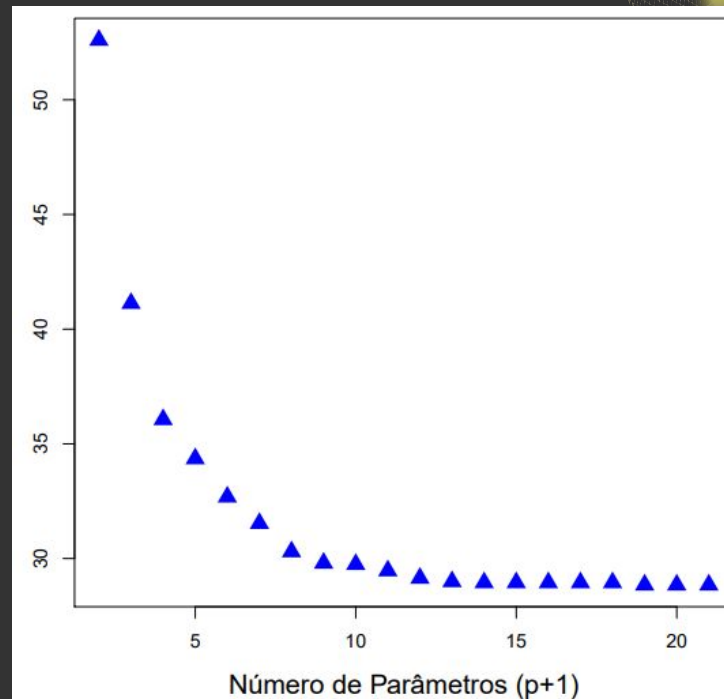




### 3. Seleção de modelos

# Seleção de modelos

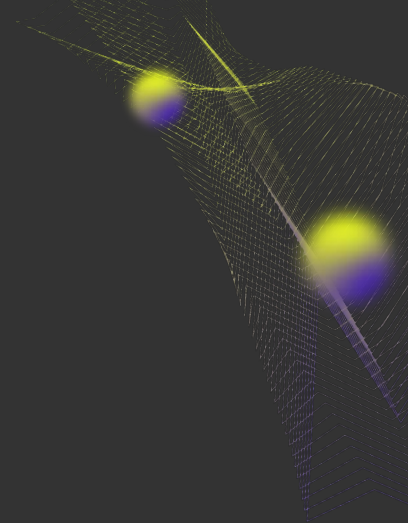
- Calcular MSE nos dados de treino para selecionar melhor modelos
- Abordagem ruim pois com mais parâmetros os modelo sempre ficam mais ajustado aos dados de treino, perdendo poder de generalização
  - Super ajustes (overfitting)





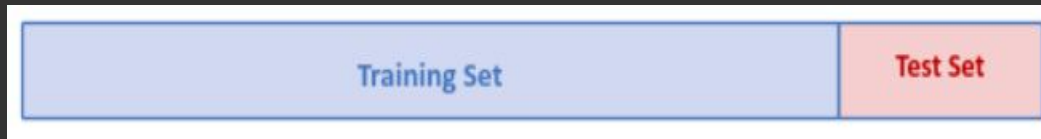
# Seleção de modelos

- Duas abordagens:
  - penalização da função perda (AIC, BIC, entre outros)
  - validação cruzada (Machine learning)
- Idéia geral: **avaliar o poder de preditivo em dados não vistos no processo de treinamento**
- Diversas abordagens: Vamos começar pela mais simples - Método holdout

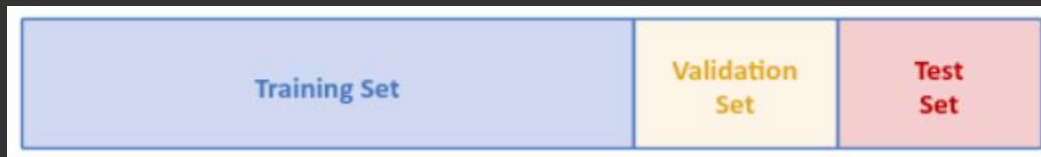


# Método holdout

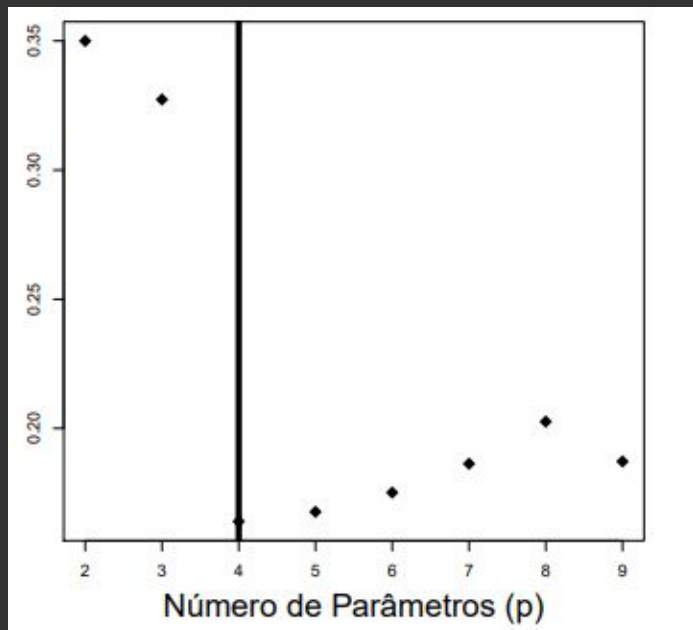
- Dividimos todo o conjunto de dados em duas partes: treinamento e teste



- Um avanço neste método é usar uma parte para validação



# Método holdout



- $p = 4$  apresentou o menor MSE no conjunto de teste.
- Idealmente fazemos a escolha de  $p$  usando dados de validação e ajustamos o modelo com treino + validação para reportar o MSE

# Método k-fold



- Mais eficiente que o método holdout
- Mais custoso computacionalmente