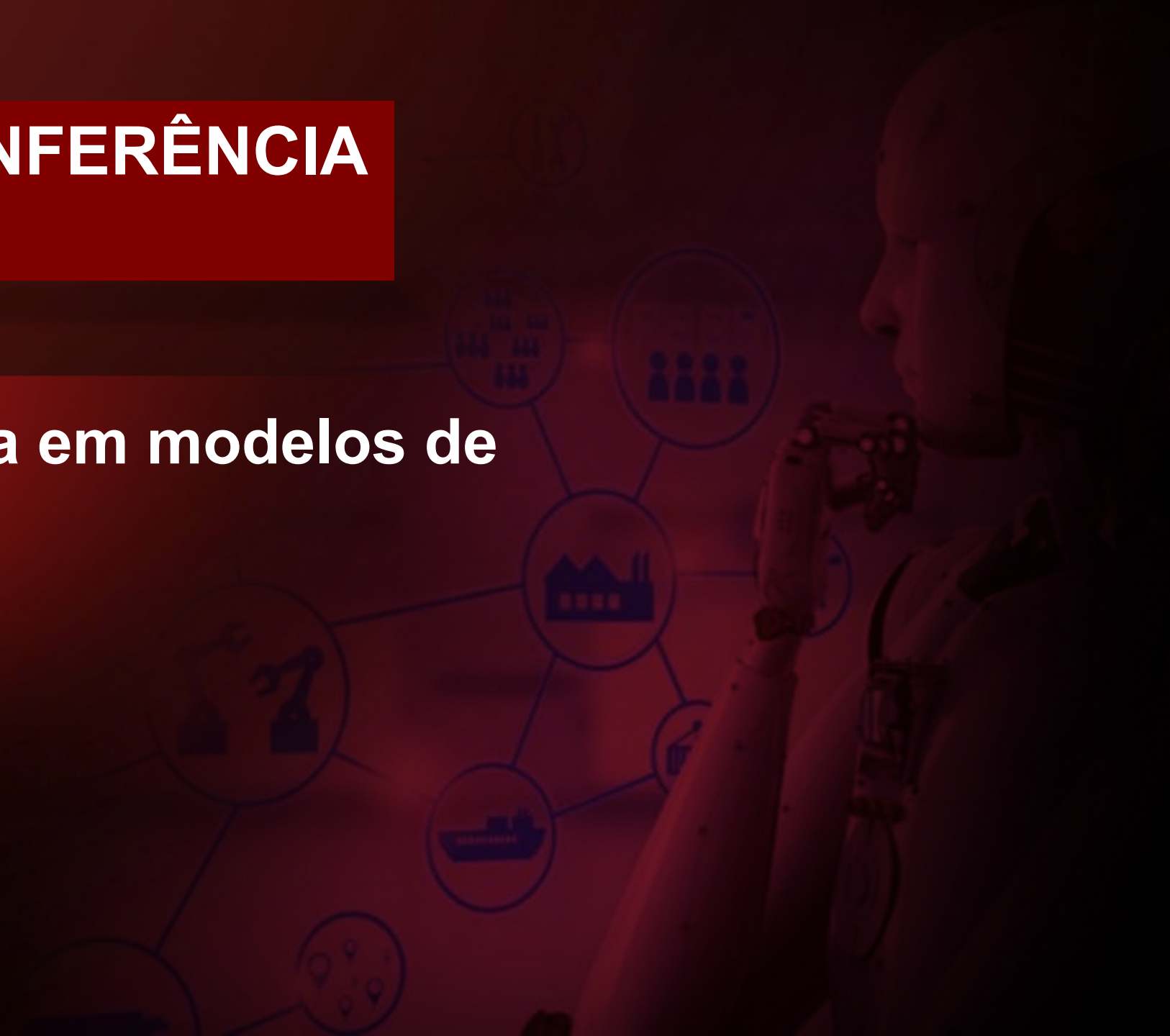


# MODELAGEM E INFERÊNCIA ESTATÍSTICA

## Critérios de escolha em modelos de regressão múltipla



# O QUE VOU ESTUDAR HOJE?

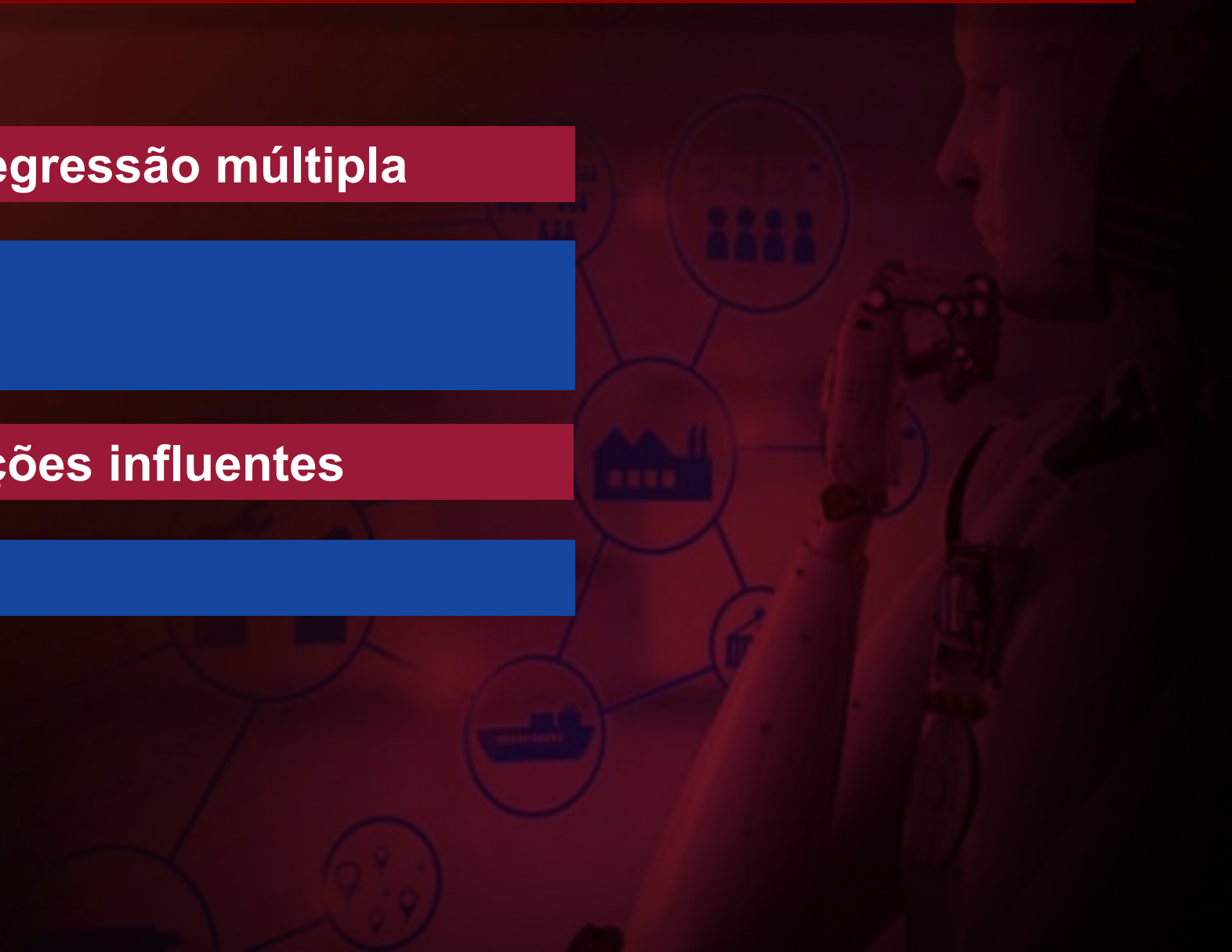
Modelos não lineares e regressão múltipla

Seleção de variáveis

- Critério  $C_k$

Identificação de observações influentes

Multicolinearidade



# MODELOS NÃO LINEARES E REGRESSÃO MÚLTIPLA

➤ Considerações sobre modelos lineares generalizados aplicados a variáveis múltiplas

- Transformação de variáveis: ex.  $Y = ax_1^{\beta_1}x_2^{\beta_2}x_3^{\beta_3}\varepsilon$
- Modelo logístico  $\frac{p(x_1, \dots, x_k)}{1-p(x_1, \dots, x_k)} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}$

➤ Gráficos de diagnóstico.

➤ Testes de hipótese.

➤ Intervalos de confiança e previsão.

# SELEÇÃO DE VARIÁVEIS

- Se alguns preditores são adicionados a um determinado modelo, o SQE não pode aumentar.
  - Um modelo maior não pode explicar menos variação do que um menor.
- SQE, geralmente, irá diminuir, talvez por uma pequena quantidade.
- O modelo que dá o maior valor  $R^2$  – deve ser o que contém todos os preditores.
- **Desejável:**
  - Um modelo envolvendo relativamente poucos preditores.
  - Que seja fácil de interpretar e usar.
  - Que explique uma quantidade relativamente grande da variação y observada.



# SELEÇÃO DE VARIÁVEIS

## CRITÉRIOS PARA K NÚMERO DE PREDITORES:

1.  $R_k^2 \rightarrow$   
aumenta se  
k aumenta.

- O ideal é trabalhar com o menor valor possível de k.
- Para esse valor pequeno de k garantir um  $R_k^2$  grande.

2.  $QME_k = SQE_k / (n - k - 1)$  quadrado médio dos erros para um modelo com k preditores.

- Quanto menos preditores maior o  $QME_k$
- Objetivo  $\rightarrow$  Modelo com o mínimo  $QME_k$
- $R_k^2(adj) = 1 - \frac{QME_k}{QMT_k}$
- $QMT_k = \frac{SQT}{n-1}$
- Avaliar  $QME_k$  ao invés de  $R_k^2$

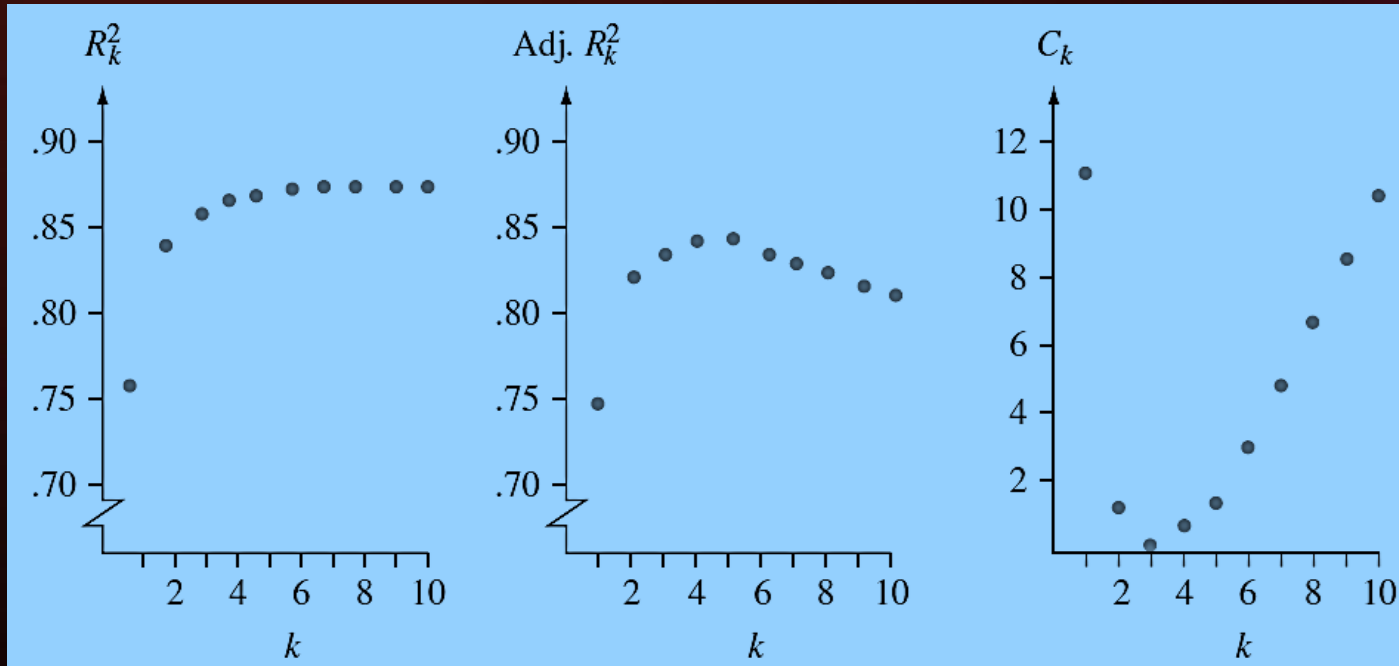
3.  $C_k = \frac{SQE_k}{s^2} + 2(k + 1) - n$

- Desejável um valor pequeno de  $C_k$

# SELEÇÃO DE VARIÁVEIS

## CRITÉRIOS PARA K NÚMERO DE PREDITORES:

1.  $R_k^2 \rightarrow$  aumenta se  $k$  aumenta.
2. Desejável um valor pequeno de  $C_k$



Fonte: (DEVORE, 2018, p.567)

# SELEÇÃO DE VARIÁVEIS

- **Eliminação retroativa (ER)** (*backward elimination*)

Exemplo: Se  $t_{out} = 2$

$$\min_{i=1,\dots,m} \left| \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \right| < t_{out}$$

Step	Predictors	t - ratio				
		1	2	3	4	5
1	1, 2, 3, 4, 5	16.0	10.8	2.9	2.8	1.8*
2	1, 2, 3, 4	15.4	10.2	3.7	2.0*	—
3	1, 2, 3	14.5	12.2	4.3*	—	—
4	1, 2	10.9	9.1*	—	—	—
5	1	4.4*	—	—	—	—

Fonte: (DEVORE, 2018, p.568)

- **Seleção progressiva (SP)** (*forward selection*)
  - Ajustar o modelo com apenas  $x_i$ , depois adiciona mais um  $x_i$

# IDENTIFICAÇÃO DE OBSERVAÇÕES INFLUENTES

Um método para identificar observações potencialmente influentes

$\hat{\beta}_i$  é uma função linear de  $y_1, y_2, \dots, y_k$

$\hat{y}$  é função linear de  $y_j$

$$\hat{y}_1 = h_{11}y_1 + h_{12}y_2 + \dots + h_{1n}y_n$$

$$\hat{y}_2 = h_{21}y_1 + h_{22}y_2 + \dots + h_{2n}y_n$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$\hat{y}_n = h_{n1}y_1 + h_{n2}y_2 + \dots + h_{nn}y_n$$

$$h_{ij} = h_{ji}$$

$$0 \leq h_{jj} \leq 1$$



# MULTICOLINEARIDADE

Definição: “Fenômeno estatístico no qual duas ou mais variáveis em um modelo de regressão são dependentes de outras variáveis de tal forma que uma pode ser predita linearmente da outra com alto grau de precisão”

(Sureka, S & Vaidya, D., 2022)

Grande

$$V(\hat{\beta}_i) = \frac{\sigma^2}{\sum_{j=1}^n (x_{ij} - \hat{x}_{ij})^2}$$

Pequeno

$R^2$  valor próximo de 1 mas  $t_i = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$  adota valor pequeno

# MULTICOLINEARIDADE

## 1. Multicolinearidade Perfeita.

Existe quando as variáveis independentes na equação predizem a relação linear perfeita.

## 2. Alta Multicolinearidade.

Relação linear entre duas ou mais variáveis independentes que não estão perfeitamente correlacionadas entre si.

## 3. Multicolinearidade Estrutural.

Causada pelo próprio pesquisador ao inserir diferentes variáveis independentes na equação.

## 4. Multicolinearidade baseada em dados.

Causada por experimentos mal planejados pelo pesquisador.

(Sureka, S & Vaidya, D., 2022)

# MODELAGEM E INFERÊNCIA ESTATÍSTICA

## Critérios de escolha em modelos de regressão múltipla

