



# GET00211 - Modelos Lineares 2

Rafael Erbisti

Instituto de Matemática e Estatística  
Universidade Federal Fluminense

Aula 11

# Ponto de corte na classificação

O modelo logístico fornece **probabilidades estimadas**  $\hat{\pi}_i$ . Para transformar em uma decisão (0/1), define-se um **ponto de corte (threshold)**:

$$\hat{y}_i = \begin{cases} 1, & \hat{\pi}_i \geq c \\ 0, & \hat{\pi}_i < c \end{cases}$$

- Valor usual:  $c = 0,5$ .
- O ponto de corte pode ser ajustado conforme o contexto (ex.: minimizar falsos negativos em saúde).
- Diferentes  $c$  levam a diferentes classificações.



# Matriz de confusão

A matriz resume o desempenho do modelo em termos de classificação:

	Classe Real = 1	Classe Real = 0
Previsto = 1	VP (Verdadeiro Positivo)	FP (Falso Positivo)
Previsto = 0	FN (Falso Negativo)	VN (Verdadeiro Negativo)

- VP = acertos de classe positiva
- VN = acertos de classe negativa
- FP = erros ao prever positivo
- FN = erros ao prever negativo



# Medidas de desempenho

A partir da matriz de confusão:

- Acurácia:  $\frac{VP + VN}{VP + VN + FP + FN}$
- Sensibilidade (Recall):  $\frac{VP}{VP + FN}$
- Especificidade:  $\frac{VN}{VN + FP}$
- Precisão:  $\frac{VP}{VP + FP}$
- F1-Score: média harmônica entre precisão e recall

Essas medidas variam de acordo com o ponto de corte escolhido.



- A **curva ROC** (Receiver Operating Characteristic) mostra a relação entre:

**Sensibilidade vs. 1 - Especificidade**

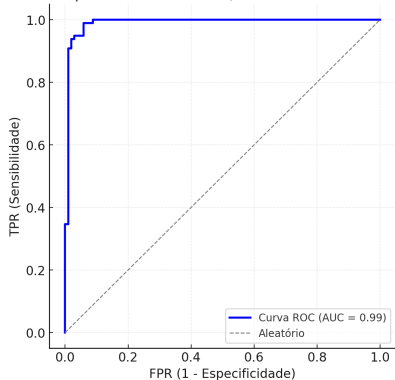
para todos os valores possíveis de ponto de corte.

- Permite visualizar o *trade-off* entre detectar positivos e evitar falsos alarmes.
- A **AUC** (Área sob a curva) mede a capacidade de discriminação do modelo:
  - $AUC = 0,5$ : **modelo aleatório**.
  - $AUC = 1$ : **modelo perfeito**.
  - Quanto maior a AUC, melhor é o modelo na separação das classes.
- A curva ROC não depende do ponto de corte específico, mas resume todos os possíveis.

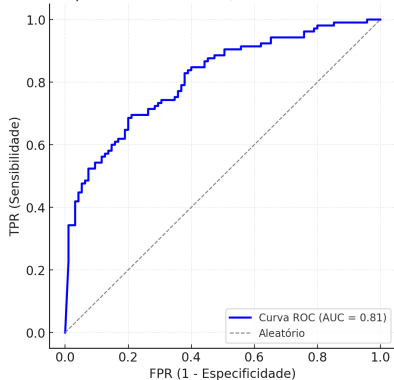


# Curva ROC e AUC

Exemplo de Curva ROC (Modelo mais acurado)



Exemplo de Curva ROC (Modelo intermediário)



**Exercício:** Considere a base `baseModelo.rds`, contendo registros de casos e óbitos por COVID-19 nos municípios do estado do Rio de Janeiro, por semana epidemiológica, no período entre 22/03/2020 a 31/12/2022 (totalizando 145 semanas). Modele a probabilidade de óbito por COVID-19 e apresente todas as etapas da análise, incluindo exploração dos dados, especificação do modelo, ajuste, avaliação de desempenho e interpretação dos resultados.

A base de dados `baseModelo.rds` e os respectivos dicionários das variáveis `dicionarioIVM.xlsx` e `dicionarioIVS.xlsx` estão disponíveis [aqui](#).

