

# Trabalho 1 - Modelos de Regressão e Aprendizado Supervisionado/Modelos Lineares Generalizados

---

POSMAC-2023.1 e 4.ano - Estatística/FCT

## Introdução

Este trabalho é uma aplicação prática dos conceitos apresentados na matéria Modelos de Regressão e Aprendizado Supervisionado/Modelos Lineares Generalizados, ministrada na POSMAC-2023.1 e no 4º ano do curso de Estatística da FCT. O exercício é baseado na parte de aplicação do livro "Modelos de Regressão com apoio computacional" de Gilberto A. Paula, professor do Instituto de Matemática e Estatística da Universidade de São Paulo.

Os dados utilizados neste trabalho foram obtidos no repositório do autor, disponível em: <https://www.ime.usp.br/~giapaula/dados.html>. O arquivo de dados `canc3.dat` contém informações sobre um estudo de caso-controle realizado no Setor de Anatomia e Patologia do Hospital Heliópolis em São Paulo, no período de 1970 a 1982.

## Objetivo

O objetivo deste trabalho é ajustar um modelo logístico com efeitos principais utilizando a biblioteca `statsmodels` em Python. A análise tem como base os dados do arquivo `canc3.dat`, referentes a um estudo de caso-controle de pacientes com processo infeccioso pulmonar.

## Metodologia

1. Importação e tratamento dos dados: carregamos o arquivo `canc3.dat` e transformamos as variáveis categóricas em fatores.
2. Ajuste do modelo logístico: utilizamos a função `GLM` da biblioteca `statsmodels` para ajustar um modelo de regressão logística com efeitos principais, considerando as variáveis idade, sexo, intensidade da célula histiócitos-linfócitos (HL) e intensidade da célula fibrose-frouxa (FF).
3. Análise dos resultados: exibimos um resumo do modelo ajustado, incluindo as estimativas dos parâmetros e seus respectivos erros padrão aproximados.

## Resultados

Os coeficientes estimados e seus respectivos erros padrão aproximados para cada variável no modelo de regressão logística ajustado foram obtidos e apresentados em uma tabela. A análise dos resultados indica que a chance de processo infeccioso maligno é maior para o sexo feminino em comparação ao masculino. Além disso, a chance de processo maligno aumenta significativamente com a idade e há indícios de que, tanto para a célula FF quanto para a célula HL, a chance de processo maligno diminui à medida que aumenta a intensidade da célula.

## Conclusão

Este trabalho apresentou uma aplicação prática dos conceitos de Modelos Lineares Generalizados utilizando a linguagem de programação Python. A análise dos dados do estudo de caso-controle permitiu obter insights sobre as variáveis que influenciam o processo infeccioso pulmonar, contribuindo para a compreensão deste fenômeno e para o desenvolvimento de estratégias de prevenção e tratamento.