

**Lista nº 4 – Estimação Não Paramétrica – Análise de Banco de Dados**

1. O banco de dados `ipeccsv` contém os dados de uma coorte de pacientes com Aids atendidos no Ipec e está detalhado abaixo. Para este estudo foi definido, como tempo de sobrevivência, o tempo entre o diagnóstico de Aids (critério CDC-1993) e o óbito.
  - a. Manipule e organize o banco de dados como se segue:
    - i. Substituir por NA o código 99 utilizado no banco para dado faltante.
    - ii. Renomeie as categorias da variável `escola` (0: sem, 1: fundam, 2: médio, 3: sup).
  - b. Amplie a análise exploratória desses dados usando o R:
    - i. Qual a idade média desses pacientes
    - ii. Qual é a razão homens:mulheres nessa coorte?
    - iii. Quantos receberam tratamento?
    - iv. Quantos foram a óbito e quantos foram censurados?
    - v. Os óbitos e as censuras se concentraram no fim do estudo ou ocorreram durante todo o período?
    - vi. Qual o tempo de sobrevivência mediano?
  - c. Existe diferença entre homens e mulheres quanto ao tempo de sobrevivência pós-diagnóstico de Aids?
  - d. Compare homens e mulheres segundo os tempos de sobrevivência mediano e percentis 75 e 90.
  - e. Existe diferença na sobrevivência dos pacientes submetidos aos diferentes tipos de tratamento (variável `tratam`)?
  - f. Existe diferença na sobrevivência de pacientes com diferentes graus de escolaridade?
  - g. Faça os gráficos de sobrevivência e risco e aplique os testes log-rank e Peto.

2. Detalhes sobre o Banco de Dados:

Dados provenientes de coortes hospitalares de pacientes portadores de HIV. A primeira coorte é constituída dos pacientes portadores de HIV atendidos entre 1986 e 2000 no Instituto de Pesquisa Clínica Evandro Chagas (Ipec/Fiocruz). Dessa coorte, obteve-se uma amostra de 193 indivíduos que foram diagnosticados como portadores de Aids (critério CDC 1993) durante o período de acompanhamento.

Variável	Descrição
id	identificação do paciente
ini	data do diagnóstico da aids (em dias)
fim	data do óbito (ou perda do paciente)
tempo	dias de sobrevivência do diagnóstico até o óbito
status	0 = censura, 1 = óbito
sexo	F = feminino, M = masculino
escola	0 = sem escolaridade, 1 = ensino fundamental, 2 = ensino médio, 3 = ensino superior
idade	idade na data de diagnóstico de Aids (20 a 68 anos)
risco	0 = homossexual masculino, 1 = usuário de drogas injetáveis, 2 = transfusão, 3 = contato sexual com HIV+, 5 = hetero com múltiplos parceiros, 6 = dois fatores de risco
acompan	acompanhamento: 0 = ambulatorial/hospital-dia, 1 = internação posterior, 2 = internação imediata
óbito	S = óbito, N = não óbito, I = ignorado
anotrat	ano do início do tratamento (1990 a 2000), 9 = sem tratamento
tratam	terapia antirretroviral: 0 = nenhum, 1 = mono, 2 = combinada, 3 = potente
doença	de apresentação: 1 = pcp, 2 = pcp pulmonar, 3 = pcp disseminada, 4 = toxoplasmose, 5 = sarcoma, 7 = outra doença, 8 = candidíase, 9 = duas doenças, 10 = herpes, 99 = definido por cd4
propcp	profilaxia para pneumocistis: 0 = sem profilaxia, 2 = primária, 3 = secundária, 4 = ambas