

# ALGEBRA LINEAR II

*Método dos Mínimos Quadrados para a Previsão de Mortes por Infarto*

Natália Carvalhinha Sacco de Lemos Basto

Universidade Federal do Rio de Janeiro

<b>Resumo</b>	<b>2</b>
<b>Introdução</b>	<b>4</b>
<b>Método</b>	<b>4</b>
<b>Resultados</b>	<b>5</b>
<b>Conclusão</b>	<b>6</b>
<b>Referências</b>	<b>6</b>

## **1. Resumo**

O trabalho busca realizar a previsão de morte por infarto de diversas pessoas. Contamos com dados ligados a idade, anemia, nível de creatina, diabetes, saída de sangue, pressão alta, nível de plaquetas no sangue, sexo biológico e fumo. A previsão foi feita a partir de um código em Python capaz de calcular, a partir de um banco de dados, os coeficientes, com o auxílio da extensão pandas, para conversão do arquivo em matriz no Python.

Para realizar tal previsão, o programa implementa os dados de forma a se adequarem a equação normal:  $Ax = b$ , em que A são os dados que auxiliam na previsão e b é a saída que define se houve ou não morte por ataque cardíaco.

Na previsão, o usuário insere os dados relacionados à saúde e o sistema criado irá retornar se houve ou não da fatídica morte.

## 2. Introdução

Esta questão foi projetada para obter uma compreensão matemática do código capaz de prever se o usuário vai morrer. Para tanto, a questão fornece diversos dados de saúde de 290 pacientes. A partir dessas informações, o arquivo é extraído do formato .xlsx para que você possa ter uma grande matriz principal em python, que pode ser fragmentada em duas matrizes: A e b, onde b é uma matriz que indica se ocorreu um óbito, apenas sim ou não, e A são todos os dados de saúde restantes. Assim, podemos ajustar as duas matrizes à equação  $Ax = b$  e temos o centro do problema, encontrando o valor do vetor x conseguimos realizar a previsão para novos dados. Este problema seria mais fácil de resolver se A fosse uma matriz quadrada, mas ela não é, portanto não há inversa, assim o caminho ideal para a resolução é por mínimos quadrados.

## 3. Método

Para a previsão foi necessário converter os dados do arquivo para uma matriz, uma lista de listas, para que dessa forma seja possível trabalhar com os dados dentro do python. Modelou-se os dados de forma que a matriz A possuísse as colunas de ‘age’ até ‘smoking’ e o vetor coluna b possuísse a coluna ‘DEATH\_EVENT’. Após a transformação de formato, o método dos mínimos quadrados já pode ser iniciado.

$$[a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ a_6 \ a_7 \ a_8 \ a_9 \ a_{10} \ a_{11}]$$

$$Ax = b$$

$$A^T A x = A^T b$$

$$x = (A^T A)^{-1} \cdot (A^T b)$$

Inicialmente, é importante definir como o método é realizado. A matriz A e o vetor coluna b já foram apresentados, e o vetor coluna x nos retorna os coeficientes. O método é utilizado para matrizes não quadradas. Isso explica o motivo de multiplicar a matriz A pela sua transposta, gerando uma matriz resultante que possui o mesmo número de linhas e colunas. Portanto, o programa realiza o cálculo da transposta e posteriormente realiza a multiplicação dessa transposta de A por A do lado esquerdo da equação e por b no lado direito. Para isolarmos o x, iremos calcular a inversa da matriz resultante da transposta de A por A. Após realizar a inversa, uma multiplicação entre as duas matrizes é feita, resultando nos nossos coeficientes.

$$b = [ a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_5 x_5 + a_6 x_6 + a_7 x_7 + a_8 x_8 + a_9 x_9 + a_{10} x_{10} + a_{11} x_{11} ]$$

Esse modelo foi aplicado para os casos de morte e não morte. Nos casos de não morte, o '0' representando o 'não' foi substituído por 2, para que os coeficientes não fossem zerados.

#### 4. Resultados

Para os casos de morte e não morte foram encontrados os coeficientes:

Caso Não Morte

$$[x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10} x_{11}] = [0,0000004379358439356531 - 0,005019185158234463 0,00000263$$

Caso Morte

$$[x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10} x_{11}] = [- 0,0000001295452302244859 - 0,0010930715921666917$$

## 5. Conclusão

O procedimento proposto neste trabalho mostrou-se eficaz para a previsão. Mesmo que os dados apresentados não constituam uma base de dados simétrica em relação às suas saídas. Portanto, a álgebra linear e os métodos computacionais aplicados neste trabalho são capazes de atingir seus objetivos com sucesso.

Para uma maior precisão o programa deveria ter sido mais aprofundado, o que exigiria maior tempo de dedicação ao sistema em questão. O trabalho ajudou muito a aprofundar o entendimento dos mínimos quadrados, principalmente quando se trabalha com grandes bancos de dados.

## 6. Referências

BOLDRINI, José Luiz *et al.* **Álgebra Linear**. 3. ed. São Paulo: Harbra, 1986.

CABRAL, Marco A. P.; GOLDFELD, Paulo. **Curso de Álgebra Linear**: fundamentos e aplicações. 3. ed. Rio de Janeiro: Instituto de Matemática, 2012.