

## XVIII EAMC – Análise de Dados com Python

### Exercício 1

**Arquivo de dados:** areas.csv

**Fonte:** MAGALHÃES, M.N.; DE LIMA, A.C.P. **Noções de Probabilidade e Estatística. 7. ed. São Paulo: EdUSP, 2007.**

Um empreendimento imobiliário consistiu na construção de dois edifícios residenciais, com apartamentos de 1 dormitório e área útil de 53 m<sup>2</sup>. Em uma primeira fase, foi construído apenas um edifício, denominado “Bloco A”. Para a conclusão do Bloco B houve uma troca de empreiteiras. Suspeita-se que seus apartamentos foram construídos com metragem diferente daquela especificada na escritura. O arquivo areas.txt contém as seguintes variáveis, obtidas por peritos de uma firma independente:

**Id:** identificação da observação no arquivo.

**Bloco:** bloco a que pertence o apartamento.

**Andar:** andar onde o apartamento está situado.

**Final:** número identificando a posição do apartamento no andar.

**Sala:** área da sala, em m<sup>2</sup>.

**Cozinha:** área da cozinha, em m<sup>2</sup>.

**Banheiro:** área do banheiro, em m<sup>2</sup>.

**Dorm:** área do dormitório, em m<sup>2</sup>.

**Rachadura:** ocorrência de rachaduras no apto

**Infiltr:** ocorrência de infiltrações no apto

1. Vamos começar explorando o tipo de empreendimento de que tratam os dados. Para isso, verifique:

- a) Qual o número total de apartamentos no empreendimento?
- b) Quantos apartamentos existem por bloco?
- c) Quantos apartamentos existem por andar?

2. Existe a suspeita de que os apartamentos tenham sido construídos com metragens diferentes. O arquivo de dados traz as áreas de cada cômodo em separado. Acrescente uma nova coluna calculando a área útil total de cada apartamento.

3. Para a área de cada cômodo e para a área útil total (isto é, para as variáveis quantitativas contínuas), faça um histograma.

4. Agora, em separado para o bloco A e o bloco B:

- a. Calcule a média e o desvio-padrão da área de cada cômodo e também da área total.
- b. Construa boxplots e compare as áreas para cada cômodo considerado.
- c. Com base em seus achados, você diria que existem diferenças nas áreas dos apartamentos dos blocos A e B? Em caso positivo, qual (is) cômodo (s) apresenta (m) o problema?

4. Agora vamos explorar os dados referentes a problemas estruturais (rachaduras e infiltrações).

- a. Construa tabelas de frequências e gráficos (barras ou pizza) para cada uma das duas variáveis.
- b. Use a tabulação cruzada para verificar se há diferença entre a ocorrência de rachaduras e infiltrações para os blocos A e B.
- c. Crie uma nova coluna, com o nome **Andar**, e nela divida os apartamentos em duas categorias: baixo (até nono andar) e alto (décimo andar em diante).
- d. Usando tabelas e / ou gráficos, estude a ocorrência de rachaduras e infiltrações em separado para cada categoria.