

# Lista 5 - Introdução a Análise de Dados

## Dataframes

Guilherme Masuko

May 2023

Para essa lista, utilizaremos uma base de dados própria do R chamada "Motor Trend Car Road Tests". Utilize os comandos `data(mtcars)` e `View(mtcars)` para importar e visualizar os dados respectivamente. Para conhecer melhor essa base de dados (entender o significado de cada coluna, por exemplo), deve-se usar o seguinte comando `?mtcars`.

### Questão 1

Encontre, utilizando código, o carro mais econômico dessa amostra (maior milhas por galão (*mpg*)).

### Questão 2

Encontre, utilizando código, o carro menos econômico dessa amostra (menor milhas por galão (*mpg*)).

### Questão 3

Dentro da categoria de carros mais potentes (digamos,  $hp > 120$ ), qual é o carro mais econômico?

### Questão 4

E dentro da categoria de carros menos potentes (digamos,  $hp \leq 120$ ), qual é o carro menos econômico?

### Questão 5

Existe um *trade-off* entre economia e potência de carros? Siga as instruções a seguir para responder essa pergunta.

- a) Crie uma função que recebe dois vetores de mesmo tamanho e retorne a covariância<sup>1</sup> (amostral) entre esses dois vetores.

---

<sup>1</sup><https://en.wikipedia.org/wiki/Covariance>

A covariância (populacional) entre duas variáveis é obtida através de

$$\mathbb{C}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Mas como nossa base não tem todos os carros do mundo, vamos utilizar um estimador para a covariância verdadeiro (populacional), esse estimador é a covariância amostral, obtida através de

$$\widehat{\mathbb{C}(X, Y)} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y)$$

onde  $\mu_X$  e  $\mu_Y$  são as médias amostrais da variáveis  $X$  e  $Y$ , respectivamente.

Obs: O aluno deve utilizar a formula do estimador para criar a função.

- b) Crie uma função que receba dois vetores de mesmo tamanho como parâmetros e retorne a correlação<sup>2</sup> (amostral) entre eles.

A correlação (populacional) entre duas variáveis é obtida através de

$$\rho_{X,Y} = \frac{\mathbb{C}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

onde  $\sigma_X$  e  $\sigma_Y$  são os desvios-padrão populacionais da variáveis  $X$  e  $Y$ , respectivamente.

E seu estimador é a correlação amostral, obtida através de

$$\widehat{\rho_{X,Y}} = \frac{\widehat{\mathbb{C}(X, Y)}}{\widehat{\sigma_X} \cdot \widehat{\sigma_Y}}$$

onde  $\widehat{\sigma_X}$  e  $\widehat{\sigma_Y}$  são os desvios-padrão amostrais da variáveis  $X$  e  $Y$ , respectivamente.

Obs: O aluno deve utilizar a formula do estimador para criar a função.

- c) Calcule a correlação entre as duas variáveis economia e potência do carro (representadas por *mpg* e *hp*, respectivamente) para responder a pergunta principal.

### Questão 6

Crie uma coluna (*wt\_kg*) que contenha o peso de libras (*wt*) convertido em quilogramas. Note que a medida da coluna *wt* equivale a 1000 libras. A fórmula de conversão é:

---

<sup>2</sup><[https://en.wikipedia.org/wiki/Pearson\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Pearson_correlation_coefficient)>

$$\text{peso em kg} = \frac{\text{peso em libras}}{2.2046}$$

**Questão 7**

Qual é o peso médio, em quilogramas, dos carros?

**Questão 8**

Qual é o peso médio, em quilogramas, dos carros automáticos ( $am = 0$ )?

**Questão 9**

Qual é o peso médio, em quilogramas, dos carros manuais ( $am = 1$ )?

**Questão 10**

Qual é a correlação entre essas duas variáveis, transmissão ( $am$ ) e peso em quilogramas ( $wt\_kg$ )? O que isso significa?

**Probabilidade**

A probabilidade de um evento  $A \subset \Omega$  ocorrer é

$$\begin{aligned} \mathbb{P}(A) &= \frac{n(A)}{n(\Omega)} \\ &= \frac{\text{número de elementos do evento}}{\text{número de elementos do espaço amostral}} \end{aligned}$$

**Questão 11**

Qual a probabilidade (amostral) de pegarmos (em nossa amostra) um carro manual mais pesado que o peso médio dos carros automáticos?

**Questão 12**

Qual a probabilidade (amostral) de pegarmos (em nossa amostra) um carro automático mais pesado que o peso médio dos carros manuais?

**Questão 13**

Qual a probabilidade (amostral) de pegarmos (em nossa amostra) um carro manual mais pesado que o carro automático mais leve?

**Questão 14**

Qual a probabilidade (amostral) de pegarmos (em nossa amostra) um carro automático mais leve que o carro manual mais pesado?