



escola
britânica de
artes criativas
& tecnologia

Profissão: Analista de dados



FUNDAMENTOS DE ESTATÍSTICA



GUIA DA AULA 3



Faça a correlação

Introdução

Coeficiente de
Correlação de
Pearson



Acompanhe aqui
os temas que
serão tratados
na videoaula



Introdução

Correlação são métricas que medem a dependência estatística entre conjuntos de dados, correlações estas que podem ser causais ou não.

- **Exemplo:**

Sobe o preço do dólar, sobe o preço de insumos importados, sobe o preço dos produtos, logo, sobe a inflação.



Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson (r_{xy}) resume a correlação linear entre dois conjuntos de dados em um único número entre -1 e 1, sendo que:

- $\text{r}_{xy} < 0$, enquanto x cresce, y decresce;
- $\text{r}_{xy} = 0$, não há relação entre x e y;
- $\text{r}_{xy} > 0$, enquanto x cresce, y cresce.

$$\text{r}_{xy} = \frac{\sum_{i=1}^n (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_{i=1}^n (x_i - x_m)^2} \sqrt{\sum_{i=1}^n (y_i - y_m)^2}}$$



Exemplo:

```

In [ ]: %%writefile
        nba.csv
        height;weight;wage
        2.01;86.2;17150000
        1.93;106.1;898310
        2.11;120.2;9881598
        1.88;85.7;15643750
        1.88;84.8;2875000
        2.11;106.1;2376840
        1.98;86.6;2625717
        2.08;104.3;37199000
        2.03;117.9;28942830
        1.83;81.6;522738
  
```

```

In [ ]: df = pd.read_csv('nba.csv', sep=';')
  
```



```
In [ ]: df.head()
```

```
In [ ]: height_array =
np.array(df['height'].to_list()) weight_array
= np.array(df['weight'].to_list()) wage_array
= np.array(df['wage'].to_list())
```

```
In [ ]: weight_array
```



- **Peso e altura**

```
In [ ]: np.corrcoef(weight_array, height_array)
```

```
In [ ]: with sns.axes_style('whitegrid'):

    grafico = sns.scatterplot(x=weight_array,
                              y=height_array) grafico.set
        title='Distribuição de Peso e Altura dos Jogadores da NBA em
              2020', xlabel='Peso (kg)',
        ylabel='Altura (m)'
    );
```



- **Peso e salário**

```
In [ ]: np.corrcoef(weight_array, wage_array)
```

```
In [ ]: with sns.axes_style('whitegrid'):

    grafico = sns.scatterplot(x=weight_array,
                              y=wage_array) grafico.set(
        title='Distribuição de Peso e Salário dos Jogadores da NBA em
        2020', xlabel='Peso (kg)',
        ylabel='Salário (USD)'
    );
```



- **Altura e salário**

```
In [ ]: np.corrcoef(height_array, wage_array)
```

```
In [ ]: with sns.axes_style('whitegrid'):

    grafico = sns.scatterplot(x=height_array,
                             y=wage_array) grafico.set(
        title='Distribuição de Altura e Salário dos Jogadores da NBA em
        2020 xlabel='Peso (kg)',
        ylabel='Salário (USD)'
    );
```

