



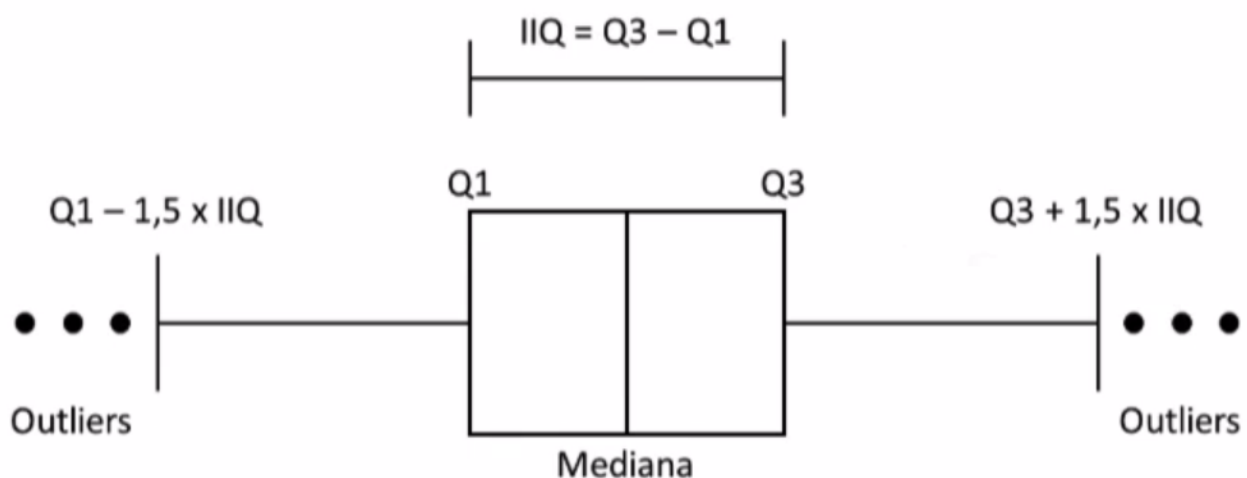
04

Usando Box plot

Transcrição

Continuaremos a análise gráfica em nosso projeto de regressão linear. Na última aula, fizemos uma visualização do histórico de consumo de cerveja ao longo do ano. Nesta aula, criaremos um *box-plot* da mesma variável para estudarmos o comportamento das estatísticas das variáveis.

Aprendemos algo similar no curso de Pandas, inclusive, utilizamos técnicas similares para excluir *outliers* do dataset. O box-plot é construído da seguinte maneira: possui uma **mediana**, divida em primeiro e segundo quartil (que corresponde aos 25%), isto é, **Q1**; temos o terceiro quartil **Q3** os 50%. Em seguida, na parte superior do gráfico, teremos o **QII**, ou intervalo interquartil, obtido por meio do resultado de **Q3 - Q1**, também utilizado para calcular os limites dos outliers.



Box-plot

Agora, seremos introduzidos à biblioteca **Seaborn**. Trata-se de uma biblioteca de visualização de dados voltados para o universo de estatística. Para mais informações consulte a documentação, facilmente encontrada on-line.

Importaremos o Seaborn escrevendo:

```
import seaborn as sns
```

[COPIAR CÓDIGO](#)

Feito isso, criaremos nossos gráficos utilizando essa biblioteca. Criaremos um box-plot utilizando a variável dependente, faremos isso escrevendo:

```
sns.boxplot()
```

[COPIAR CÓDIGO](#)

No notebook você também encontrará o link da documentação com alguns exemplos para explorar. Em seguida, escreveremos os dados a serem plotados:

```
sns.boxplot(data=dados['consumo'])
```

[COPIAR CÓDIGO](#)

Podemos, inclusive, modificar a orientação do gráfico adicionando `orient='h'` ou `orient='v'`. Outra configuração possível é o `width`, em que podemos definir a largura do boxplot.

Uma vez que a distribuição de dados pareça estar correta, podemos trabalhar na formatação. Uma das boas práticas é atribuir o gráfico à uma variável, que chamaremos de `ax`. Em seguida, podemos trabalhar a formatação ponto a ponto.

```
ax = sns.boxplot(data=dados['consumo'], orient='v', wid
```

[COPIAR CÓDIGO](#)

No Seaborn, cada tipo de gráfico possui sua própria configuração, então muitas vezes não temos uma padronização de recursos. Aprenderemos detalhes de cada um dos tipos. Para modificarmos seu tamanho, utilizaremos o `figure.set_size_inches`. Configuraremos para `(12, 6)`. Dessa maneira, teremos uma figura um pouco maior. Em seguida, inseriremos o título `Consumo de Cerveja`. Lembrando que o Seaborn foi construído sobre o Matplotlib, então todos os outputs possuem essa referência.