

Conjuntos de treino e teste

A técnica de separação aleatória de um *dataset* em dois conjuntos distintos (conjuntos de treino e teste), estimar o modelo com os dados de um destes conjuntos (conjunto de treino) e posteriormente realizar estimações com os dados do segundo conjunto (conjunto de teste) é uma técnica bastante utilizada em *data science*, para confirmar a eficácia do modelo estimado.

O **scikit-learn** disponibiliza uma função que faz essa separação, basta que informemos os dados de nosso modelo e qual o tamanho desejado dos conjuntos de treino e teste. Seria da seguinte forma o código:

```
train_test_split(X, y, test_size=0.3)
```

Onde **x** é o nosso conjunto de variáveis explicativas, **y** a nossa variável dependente e **test_size** o percentual da base que desejamos separar para testes, no caso acima, 30%.

Esta função retorna, para o conjunto de variáveis explicativas (**x**), um conjunto de treino e outro de teste (**x_train** e **x_test**) e para a variável dependente (**y**), um conjunto de treino e outro para teste (**y_train** e **y_test**). Assinale a alternativa que mostra a ordem correta de retorno desta função.

Selecione uma alternativa

A **x_train** , **x_test** , **y_train** , **y_test**

B **x_test** , **x_train** , **y_test** , **y_train**

C **y_test** , **y_train** , **x_test** , **x_train**

D **y_train** , **y_test** , **x_train** , **x_test**