



50%

ATIVIDADES
3 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



14.3k xp

a

03

Teste de hipótese

Transcrição

Os grupos que pediram e que não pediram sobremesa se comportam de formas diferentes quando se trata do valor de gorjeta. Utilizamos os dados de uma amostra, pois não podemos verificar o comportamento de todas as pessoas do mundo e colocar os dados dentro de nosso plot.

Porém, precisaremos pensar se os clientes que **futuramente** virão ao nosso restaurante se comportarão da mesma maneira ou não. Vimos antes que, adicionando `porcentagem`, as diferenças entre os grupos fica bem mais visível.

Neste passo, descobriremos se a população geral que consome a sobremesa e a que não consome se comportam igual ou diferentemente da amostra através de **testes de hipótese**.

Organizaremos nosso projeto adicionando um novo título e digitando "##Teste de hipótese". Como primeiro passo, começaremos com a **hipótese nula**; adicionaremos mais uma célula textual para termos um escrito em negrito na parte superior e escreveremos `**H^{null}**`.

Nossa hipótese nula pressupõe que a distribuição da gorjeta é **igual** nos dois grupos. Na mesma célula, adicionamos a descrição da hipótese: "A distribuição da taxa da gorjeta é a mesma nos dois grupos". Lembramos que este conteúdo foi visto no curso de **Estatística II**.



50%

ATIVIDADES
3 de 6DISCORD
ALURAFÓRUM DO
CURSOVOLTAR
PARA
DASHBOARDMODO
NOTURNO

14.3k xp

a

Em seguida, copiaremos e colaremos a sentença na sequência, substituindo `null` por `alt` para nossa **hipótese alternativa**. A descrição é a **oposta** da hipótese anterior: "A distribuição da taxa da gorjeta **não** é a mesma nos dois grupos".

Para realizarmos nosso teste de hipóteses, importaremos a biblioteca **Ranksums** a partir de `scipy.stats`.

```
from scipy.stats import ranksums
```

COPIAR CÓDIGO

Executando com "Shift + Enter", a biblioteca já estará presente em nosso projeto.

Agora, traremos o valor da porcentagem de todas as pessoas que pediram a sobremesa a partir da montagem de uma `query`. Na célula seguinte, escreveremos `gorjetas.query()` recebendo `sobremesa == 'Sim'` entre aspas. Lembramos que o `Sim` e `Não` devem estar escritos exatamente como implementamos na base de dados. Em seguida, adicionamos `porcentagem` que queremos verificar.

```
gorjetas.query("sobremesa == 'Sim']").porcentagem
```

COPIAR CÓDIGO

O sistema retorna uma lista com todos os valores relativos ao nosso comando. Para deixarmos mais legível, atribuiremos todo o resultado da `query` em uma variável `sobremesa`.

```
sobremesa = gorjetas.query("sobremesa == 'Sim']").porcentagem
```

COPIAR CÓDIGO



50%

ATIVIDADES
3 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



14.3k xp

a

Faremos a mesma operação para identificar o resultado nos casos dos clientes que não pediram sobremesa, armazenando em uma variável `sem_sobremesa` e substituindo 'Sim' por 'Não' .

```
sobremesa = gorjetas.query("sobremesa == 'Sim']").porcentagem  
sem_sobremesa = gorjetas.query("sobremesa == 'Não']").porcentagem
```

COPIAR CÓDIGO

Agora que já temos os dois grupos para nossa análise, verificaremos as hipóteses através de `ranksums()` , passando os dois grupos.

```
ranksums(sobremesa, sem_sobremesa)
```

COPIAR CÓDIGO

Por enquanto, a resposta que nos importará é o `pvalue` de arredondados `0.52` , o que significará que, apesar de os dois grupos terem dado gorjetas de valores diferentes entre si, a diferença é **insignificante** do ponto de vista matemático. Ou seja, não poderemos dizer que a população geral será diferente também.

Cientes disso, continuaremos somente com nossa hipótese nula, não aceitando mais a hipótese alternativa. Para aceitarmos esta última, o valor de `pvalue` deveria ser igual ou menor do que `0.05` . Como nosso resultado foi maior do que isso, poderemos descartar esta hipótese alternativa.

Para melhorarmos a exibição em nosso relatório, atribuiremos este valor em uma variável `r` e geraremos um registro com `print()` , recebendo 'o valor do p-value é {}' seguido de



50%

ATIVIDADES
3 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



14.3k xp

a

```
format(r.pvalue) .
```

```
r = ranksums(sobremesa, sem_sobremesa)
print('O valor do p-value é {}'.format(r.pvalue))
```

COPIAR CÓDIGO

Feito isso, teremos o valor de forma visivelmente mais agradável. Copiaremos e colaremos abaixo deste resultado a descrição da hipótese nula em uma célula textual para deixar bem claro.