**03 - Testando uma nova linha de shampoos**

Um fabricante de cosméticos afirma que a adição de um novo composto químico em sua linha de shampoos consegue promover em mais de 2 centímetros o crescimento dos fios de cabelo em um período de 60 dias. Duas amostras de pessoas foram selecionadas e testadas, uma utilizando o shampoo novo (com o composto) e a outra com o shampoo antigo (sem o composto).

Os resultados (crescimento dos fios de cabelo em centímetros) podem ser verificados na tabela abaixo:

| **Shampoo Novo** | **Shampoo Antigo** |
| --- | --- |
| 3,4 | 0,3 |
| 4,9 | 1,2 |
| 2,8 | 1,2 |
| 5,5 | 1,7 |
| 3,7 | 1,1 |
| 2,5 | 0,6 |
| 4,3 | 1,2 |
| 4,6 | 1,5 |
| 3,7 | 0,5 |
| 3,4 | 0,7 |

**Dados da tabela acima:**

shampoo\_Novo = pd.Series([3.4, 4.9, 2.8, 5.5, 3.7, 2.5, 4.3, 4.6, 3.7, 3.4])

shampoo\_Antigo = pd.Series([0.3, 1.2, 1.2, 1.7, 1.1, 0.6, 1.2, 1.5, 0.5, 0.7])COPIAR CÓDIGO

Assumindo um nível de confiança de 95% e considerando que as populações se distribuem como uma normal, podemos acreditar na afirmação do fabricante do shampoo? Assinale a alternativa que apresenta a estatística de teste e a decisão correta do teste.

**Um pouco mais de teoria:** como se trata de um problema um pouco diferente do apresentado em nossa aula, vamos esclarecer alguns pontos para ajudar na solução:

**1)** Em testes entre duas amostras, quando realizamos a escolha da distribuição amostral adequada (passo 2) e perguntamos se n ≥ 30, temos que considerar que n = n1 + n2, onde n1 é o tamanho da primeira amostra e n2 o tamanho da segunda;

**2)** Quando n1 + n2 ≥ 30, utilizamos z (normal), e quando n1 + n2 < 30, σ não for conhecido e as populações forem normalmente distribuídas, utilizamos t (*t-Student*);

**3)** Quando utilizamos a *tabela t de Student*, em teste de duas amostras, os graus de liberdade são obtidos da seguinte forma: n1 + n2 - 2;

**4)** Quando o problema nos pergunta se podemos acreditar na afirmação do fabricante, está nos indicando o que devemos testar, ou seja, a nossa hipótese alternativa (H1), que no caso é:



Onde:

* **μ1** = Crescimento médio dos cabelos com o uso do novo shampoo
* **μ2** = Crescimento médio dos cabelos com o uso do shampoo antigo.

**5)** Em nosso próximo vídeo, utilizaremos o ztest\_ind() para solucionar problemas como este. Um teste similar ao ztest\_ind(), que utiliza a distribuição *t de Student*, é o ttest\_ind(). [Aqui, você será redirecionado para a documentação](https://www.statsmodels.org/stable/generated/statsmodels.stats.weightstats.CompareMeans.ttest_ind.html). Observe que o ttest\_ind() retorna a estatística de teste, o p-valor e também os graus de liberdade.

Parte superior do formulário

* Alternativa correta



* + t = 2,6646
  + Rejeitar H0, ou seja, a alegação do fabricante é verdadeira

Alternativa correta! Ao nível de confiança de 95%, rejeitamos a hipótese nula e concluímos que a alegação do fabricante é estatisticamente significativa. Segue duas opções de solução:

from scipy.stats import t as t\_student

import numpy as np

import pandas as pd

shampoo\_Novo = pd.Series([3.4, 4.9, 2.8, 5.5, 3.7, 2.5, 4.3, 4.6, 3.7, 3.4])

shampoo\_Antigo = pd.Series([0.3, 1.2, 1.2, 1.7, 1.1, 0.6, 1.2, 1.5, 0.5, 0.7])

media\_A = shampoo\_Novo.mean()

desvio\_padrao\_A = shampoo\_Novo.std()

media\_B = shampoo\_Antigo.mean()

desvio\_padrao\_B = shampoo\_Antigo.std()

significancia = 0.05

confianca = 1 - significancia

n\_A = len(shampoo\_Novo)

n\_B = len(shampoo\_Antigo)

D\_0 = 2

graus\_de\_liberdade = n\_A + n\_B - 2

t\_alpha = t\_student.ppf(confianca, graus\_de\_liberdade)

numerador = (media\_A - media\_B) - D\_0

denominador = np.sqrt((desvio\_padrao\_A \*\* 2 / n\_A) + (desvio\_padrao\_B \*\* 2 / n\_B))

t = numerador / denominador

print('t =', round(t, 4))

if(t >= t\_alpha):

print('Rejeitar H0')

**else:**

print('Aceitar H0')COPIAR CÓDIGO

Ou, como veremos no próximo vídeo:

from statsmodels.stats.weightstats import DescrStatsW

shampoo\_Novo = pd.Series([3.4, 4.9, 2.8, 5.5, 3.7, 2.5, 4.3, 4.6, 3.7, 3.4])

shampoo\_Antigo = pd.Series([0.3, 1.2, 1.2, 1.7, 1.1, 0.6, 1.2, 1.5, 0.5, 0.7])

significancia = 0.05

test\_A = DescrStatsW(shampoo\_Novo)

test\_B = DescrStatsW(shampoo\_Antigo)

test = test\_A.get\_compare(test\_B)

t, p\_valor, df = test.ttest\_ind(alternative='larger', value=2)

print('t =', round(t, 4))

print('p-valor =', round(p\_valor, 4))

print('graus de liberdade =', df)

if(p\_valor <= significancia):

print('Rejeitar H0')

else:

print('Aceitar H0')

Parte inferior do formulário

*  t = 1,7341
* Rejeitar H0, ou seja, a alegação do fabricante é verdadeira

Alternativa errada! O valor de t desta alternativa é, na verdade, o valor de tα. Este valor é o ponto de separação entre áreas de aceitação e rejeição da hipótese nula.





* t = 8,7206
* Rejeitar H0, ou seja, a alegação do fabricante é verdadeira

Alternativa errada! Observe que neste teste, estamos assumindo que a hipótese nula é  e não que .





* t = 2,6646
* Aceitar H0, ou seja, a alegação do fabricante não é verdadeira

Alternativa errada! Verifique se as hipóteses do teste foram formuladas corretamente o se a regra de decisão foi aplicada de forma correta.