

AAG04

- Desenvolva em Python um script que execute os seguintes passos:
 1. Escolha um contexto (tema) e colete ou gere pequenas amostras comparáveis, com $n < 30$.
 2. Execute Kolmogorov-Smirnov para verificar se as amostras seguem distribuição Normal (p -value de pelo menos 0.8).
 3. Repita de 1 e 2 até encontrar pelo menos 2 amostras Normais.
 4. Verifique se as médias das amostras possuem diferença significativa (Teste-T).
 5. Calcule o intervalo de confiança para as médias para níveis de confiança de 90%, 95% e 99%.
 6. Apresente conclusões sobre os resultados obtidos nas etapas 2, 4 e 5.
- Regras:
 1. Funções prontas de bibliotecas Python **DEVEM** ser usadas ao máximo possível.
 2. Código e resultados devem ser explicados em Markdown com comandos LaTeX.
 3. Os formatos de entrega devem ser `.pdf` e `.ipynb` (código fonte + markdowns).
 4. Os dados devem ser entregues em anexo.

1. Contexto e Amostras

Vamos gerar amostras aleatórias de um tema fictício, como tempos de resposta de dois servidores. Utilizaremos numpy para gerar os dados.

```
In [88]: import numpy as np
import scipy.stats as stats

# Geração de amostras aleatórias com tamanho < 30
np.random.seed(10) # Para reprodutibilidade
sample1 = np.random.normal(loc=50, scale=5, size=25) # Média 50, desvio padrão 5
sample2 = np.random.normal(loc=55, scale=5, size=25) # Média 55, desvio padrão 5
# print("Amostra 1:", sample1)
# print("Amostra 2:", sample1)
```

2. Teste de Kolmogorov-Smirnov para normalidade

Se ambos os p -value forem maiores que 0.8 podemos seguir para o próximo passo, caso contrário é só gerar novas amostras

```
In [89]: # Teste de normalidade para as amostras
ks1 = stats.kstest(sample1, 'norm', args=(np.mean(sample1), np.std(sample1)))
ks2 = stats.kstest(sample2, 'norm', args=(np.mean(sample2), np.std(sample2)))

# Verificação do p-value
print("Amostra 1: p-value =", ks1.pvalue)
print("Amostra 2: p-value =", ks2.pvalue)
```

Amostra 1: p-value = 0.657946069111738

Amostra 2: p-value = 0.4913557472750222

3. Repetição até encontrar amostras normais

Como os p-values das amostras anteriores eram inferiores a 0,8, gerar novas amostras até que ambos sejam maiores ou iguais a 0.8.

```
In [90]: while ks1.pvalue < 0.8 or ks2.pvalue < 0.8:
        sample1 = np.random.normal(loc=50, scale=5, size=25)
        sample2 = np.random.normal(loc=55, scale=5, size=25)
        ks1 = stats.kstest(sample1, 'norm', args=(np.mean(sample1), np.std(sample1)))
        ks2 = stats.kstest(sample2, 'norm', args=(np.mean(sample2), np.std(sample2)))

        print("Amostra 1: p-value =", ks1.pvalue)
        print("Amostra 2: p-value =", ks2.pvalue)
        print("Novas amostras encontradas com p-value >= 0.8")
```

Amostra 1: p-value = 0.8778909693732064

Amostra 2: p-value = 0.9165868890127113

Novas amostras encontradas com p-value >= 0.8

4. Teste-T para diferença de médias

Usamos o teste t para verificar se as médias são significativamente diferentes.

```
In [91]: t_test = stats.ttest_ind(sample1, sample2)
        print("Teste-T: Estatística =", t_test.statistic, ", p-value =", t_test.pvalue)
```

Teste-T: Estatística = -3.3492415493090517 , p-value = 0.001584140690012504

5. Cálculo dos intervalos de confiança

Calculando os intervalos de confiança de 90%, 95% e 99% para as médias.

```
In [102... confidence_levels = [0.90, 0.95, 0.99]

def confidence_interval(data, confidence):
    n = len(data)
    mean = np.mean(data)
    sem = stats.sem(data) # Erro padrão
    margin = sem * stats.t.ppf((1 + confidence) / 2, n - 1)
    return mean - margin, mean + margin

for confidence in confidence_levels:
    ci1 = confidence_interval(sample1, confidence)
    ci2 = confidence_interval(sample2, confidence)
    print(f"Intervalo de confiança ({int(confidence*100)}%) \
          \n Amostra 1: ({ci1[0]:.2f}, {ci1[1]:.2f}) \
          \n Amostra 2: ({ci2[0]:.2f}, {ci2[1]:.2f})")
```

Intervalo de confiança (90%)

Amostra 1: (47.79, 51.21)

Amostra 2: (52.50, 55.77)

Intervalo de confiança (95%)

Amostra 1: (47.44, 51.56)

Amostra 2: (52.16, 56.11)

Intervalo de confiança (99%)

Amostra 1: (46.70, 52.30)

Amostra 2: (51.46, 56.81)

Conclusões

1. Normalidade das Amostras:

As amostras 1 e 2 passaram no teste de Kolmogorov-Smirnov com p -values de 0.8779 e 0.9166,

respectivamente, ambos superiores ao critério de $p\text{-value} \geq 0.8$. Portanto, podemos considerar que ambas seguem uma distribuição normal.

2. Teste de Diferença de Médias (Teste-T):

O teste-T indicou que há uma diferença estatisticamente significativa entre as médias das duas amostras, com um $p\text{-value}$ de 0.0016, que é muito menor que o nível de significância padrão de 0.05. A estatística t calculada foi -3.3492, reforçando essa diferença significativa.

3. Intervalos de Confiança:

- **90% de confiança:**

- Amostra 1: (47.79, 51.21)
- Amostra 2: (52.50, 55.77)

Não há sobreposição entre os intervalos, indicando uma clara separação entre as médias das duas amostras.

- **95% de confiança:**

- Amostra 1: (47.44, 51.56)
- Amostra 2: (52.16, 56.11)

Também não há sobreposição, confirmando a diferença das médias.

- **99% de confiança:**

- Amostra 1: (46.70, 52.30)
- Amostra 2: (51.46, 56.81)

Mesmo para um nível de confiança mais rigoroso, os intervalos continuam separados, reforçando a evidência de diferença entre as médias.

4. Conclusão Geral:

Com base nos resultados do teste-T e nos intervalos de confiança, concluímos que há uma diferença estatisticamente significativa entre as médias das amostras 1 e 2. Essa diferença é consistente mesmo em níveis de confiança elevados (até 99%). Portanto, as duas populações podem ser consideradas distintas em termos de suas médias.