

# MQ04

November 10, 2019

## 1 Trabalho 04 - Métodos Quantitativos

Mayara Amanda da Silva<sup>1</sup> - Redes de Computadores, Pedro Henrique Ventura<sup>1</sup> - Sistemas Multimídia

1 - Alunos da Pós Graduação em Ciência da Computação (2019). Terceiro Trimestre

**Tarefa AAG04** Desenvolva em Python um script que execute os seguintes passos:

### 1.1 Bibliotecas

```
[74]: #Manipulação de dados, bases, calculos e geração de gráficos
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import scipy
import math
```

## 2 Gerando as Bases

1 - Escolha um contexto (tema) e colete ou gere pequenas amostras comparáveis, com  $n < 30$

As amostras geradas visam mostrar as oscilações no preço das passagens interestaduais em um período de 24 meses. Vale salientar que as amostras pegam apenas 10 oscilações do preço, considerando que existiram mais de 10 oscilações nesse período de tempo.

**Amostra 1** A amostra 1 refere-se as alterações nos valores de Juiz de Fora x Três Rios

```
[81]: np.random.seed(21938)
a = stats.t.rvs(100,size=10)
print (a)
```

```
[-0.41015632  0.19039594  0.19830289 -0.61920216  1.82159095  1.31588846
 -1.04636732  0.88124015 -0.26289899 -0.88082029]
```

```
[82]: amostra1 = [-0.41, 0.22, 0.19, -0.61, 1.82, 1.31,
                -1.04, 0.88, -0.26, -0.88]
```

**Amostra 2** A amostra 2 é referente aos valores de Três Rios x Juiz de Fora.

```
[83]: np.random.seed(51933)
      b = stats.t.rvs(100,size=10)
      print (b)
```

```
[-0.56445691  0.65636964 -0.30369643  0.55985072 -2.50393397  0.11675582
  0.92656235 -0.92966121 -0.15966939 -0.2745982 ]
```

```
[84]: amostra2 = [-0.56, 0.65, -0.30, 0.55, -2.50, 0.11,
                0.92, -0.92, -0.15, -0.27]
```

### 3 Validando com Kolmogorov-Smirnov

#### 2 - Execute Kolmogorov-Smirnov para verificar se as amostras seguem distribuição Normal (p-value de pelo menos 0.8)

Aplicamos o Teste de Kolmogorov-Smirnov para verificar se as amostras geradas vem de população com distribuição específica, que no exercício deve ser distribuição normal. Esse teste serve para validar as amostras antes de ser efetuado do T-Test

#### 3 - Repita de 1 e 2 até encontrar pelo menos 2 amostras Normais

Para obtermos o resultado esperado, repetimos esse processo até que as duas amostras geradas tivessem um pvalue acima de 0.8, sendo consideradas amostras normais.

```
[61]: d1 = stats.kstest(amostra1,'norm')
      print ('\n',d1)

      d2 = stats.kstest(amostra2,'norm')
      print ('\n',d2)
```

```
KstestResult(statistic=0.1491699503309814, pvalue=0.9792246775799587)
```

```
KstestResult(statistic=0.17878637961437172, pvalue=0.9065546008698484)
```

### 4 T-Test

#### 4 - Verifique se as médias das amostras possuem diferença significativa (Teste-T)

O T-Test permite comparar valores médios das duas amostras geradas e, com isso, determinar se elas vieram da mesma população. Como nossas amostras são de um mesmo grupo (referentes a passagens de ida e volta de uma cidade pra outra), é possível afirmar que essas amostras devem ter uma média e desvio padrão ligeiramente diferentes, como mostra nosso resultado.

```
[89]: print('==== Resultado ====')

t, p = stats.ttest_ind(a,b)
print('t:',t, '\np:',p)
```

```
==== Resultado ====
t: 0.8456155531053122
p: 0.408867966007028
```

Calculamos o valor de **t** usando a fórmula acima, sendo que, para cada **t**, existe um **pvalue** associado.

Sabendo que, quanto menor o valor de **p**, maior a significância estatística, podemos afirmar que: como o valor de **p** é 0.40, temos que existe 40% de chance da diferença das nossas amostras serem ao acaso.

## 5 Calculando Intervalos de Confiança

**5 - Calcule o intervalo de confiança para as médias para níveis de confiança de 90%, 95% e 99%.**

O objetivo de calcular intervalo de confiança para os níveis de 90%, 95% e 99% significa que 90%, 95% e 99% dos intervalos de confiança construídos a partir das nossas amostras, contêm um parâmetro real contido nas amostras.

Se um teste de hipótese for realizado, o nível de confiança é o complemento do nível de significância. Com isso, temos que:

- O intervalo de confiança de 90% reflete um nível de significância de 0,10.
- O intervalo de confiança de 95% reflete um nível de significância de 0,05.
- O intervalo de confiança de 99% reflete um nível de significância de 0,01.

### Amostra 1

```
[95]: #calcular a matriz média
print('\n==== Amostra 1 ====')
media = a.mean(axis=0)
print('\nMédia: {}'.format(media))

## 90%
#Desvio Padrão
dp=0.9
desvio_padrao = np.std(a)
conf_int = scipy.stats.norm.interval(dp, loc=media, scale=desvio_padrao)
print("\nIntervalo de confiança",dp,": ")
print(conf_int)

dp=0.95
conf_int = scipy.stats.norm.interval(dp, loc=media, scale=desvio_padrao)
print("\nIntervalo de confiança",dp,": ")
print(conf_int)
```

```

print(conf_int)

dp=0.99
conf_int = scipy.stats.norm.interval(dp, loc=media, scale=desvio_padrao)
print("\nIntervalo de confiança",dp," : ")
print(conf_int)

```

==== Amostra 1 ====

Média: 0.1187973321794781

Intervalo de confiança 0.9 :  
 (-1.3754208012451696, 1.6130154656041253)

Intervalo de confiança 0.95 :  
 (-1.6616733909126356, 1.899268055271592)

Intervalo de confiança 0.99 :  
 (-2.2211378395006336, 2.4587325038595895)

## Amostra 2

```

[94]: #calcular a matriz média
print('\n==== Amostra 2 ====')
media = b.mean(axis=0)
print('\nMédia: {}'.format(media))

## 90%
#Desvio Padrão
dp=0.9
desvio_padrao = np.std(b)
conf_int = scipy.stats.norm.interval(dp, loc=media, scale=desvio_padrao)
print("\nIntervalo de confiança: ",dp," : ")
print(conf_int)

dp=0.95
conf_int = scipy.stats.norm.interval(dp, loc=media, scale=desvio_padrao)
print("\nIntervalo de confiança: ",dp," : ")
print(conf_int)

dp=0.99
conf_int = scipy.stats.norm.interval(dp, loc=media, scale=desvio_padrao)
print("\nIntervalo de confiança",dp," : ")
print(conf_int)

```

==== Conjunto B ====

Média: -0.2476477570217535

Intervalo de confiança: 0.9 :  
(-1.777344514535653, 1.2820490004921454)

Intervalo de confiança: 0.95 :  
(-2.0703938682071206, 1.5750983541636137)

Intervalo de confiança 0.99 :  
(-2.643142206488828, 2.147846692445321)

Como **conclusão** dos intervalos de confiança, é possível afirmar que: nas duas amostras geradas no exercício, temos existe 90%, 95% e 99% de “chance” do intervalo conter o verdadeiro valor da média populacional calculada acima.