

Probabilidade-e-Estatistica (/github/danielesantiago/Probabilidade-e-Estatistica/tree/c989c9edef5ca0c944dca1e7c4d3690232a9a78d)

/

Atividade 4.ipynb (/github/danielesantiago/Probabilidade-e-Estatistica/tree/c989c9edef5ca0c944dca1e7c4d3690232a9a78d/Atividade 4.ipynb)

Probabilidade e Estatística - Atividade 4 e 5



PARTE A

```
In [359]: # importar bibliotecas necessárias
import pandas as pd
import matplotlib.pyplot as plt
import statistics as std
from scipy import stats
import numpy as np
import random
import math
```

```
In [360]: # importar amostra simples e exibir as 5 primeiras entradas
df = pd.read_csv("amostra_simples_ok.csv", sep = ';')
df.IMC_am = df.IMC_am.apply(lambda x: x.replace(',', '.'))
df.IMC_am = df.IMC_am.astype(float)
df.head()
```

Out[360]:

	Sexo_am	Idade_am	Peso_am	Altura_am	IES_am	IMC_am	ClassIMC_am
0	F	20	65	155	UFAC	27.1	ob
1	F	20	35	171	UFSCAR	12.0	mg
2	M	21	41	155	UFAC	17.1	mg
3	M	22	56	173	UFAC	18.7	ad
4	F	18	41	150	UFAC	18.2	mg

a) Obter o I.C. 95% para a média do PESO e testar a hipótese $H_0 : \mu = 64$ vs $H_1 : \mu \neq 64$ (nível de significância $\alpha = 5\%$) para a variável PESO. Observe que o valor 64 equivale à mediana populacional, ou seja, o teste é para saber se a média é igual à mediana.

```
In [361]: # média amostral
media = round(df.Peso_am.mean(),2)

# variância amostral
variancia = round(std.variance(df.Peso_am),2)

print("Média amostral: ", media, "\nVariância Amostral: ", variancia)
```

Média amostral: 67.59
Variância Amostral: 273.6

Hipótese a ser testada -> $H_0 : \mu = 64$
Distribuição t de student = $n - 1 = 54 - 1 = 53$ g.l.

```
In [362]: t = round(stats.t.ppf(1-0.025, 53),4)
print ("Valor na t de student: ", t)
```

Valor na t de student: 2.0057

```
In [363]: # Valor limite inferior
vli = 64 - t * math.sqrt(variancia)/math.sqrt(54)
vli
```

Out[363]: 59.48531825233864

```
In [364]: # Valor limite superior
vls = 64 + t * math.sqrt(variancia)/math.sqrt(54)
vls
```

Out[364]: 68.51468174766137

A média amostral é 67.59 e portanto não está na região de rejeição e a hipótese nula não é rejeitada ao nível de 5% de significância.

Se compararmos os dados populacionais, a média é 21.64 e a mediana é 21, onde há realmente um valor semelhante.

b) Obter o I.C. 95% para a média de ALTURA e testar a hipótese $H_0 : \mu = 166$ vs $H_1 : \mu > 166$, com nível de significância igual a 5%, para a variável ALTURA.

```
In [365]: # média amostral
media = round(df.Altura_am.mean(),2)

# variância amostral
variancia = round(std.variance(df.Altura_am),2)

print("Média amostral: ", media, "\nVariância Amostral: ", variancia)
```

Média amostral: 167.44
Variância Amostral: 107.95

Hipótese a ser testada -> $H_0 : \mu = 166$
Distribuição t de student = $n - 1 = 54 - 1 = 53$ g.l.

```
In [366]: t = round(stats.t.ppf(1-0.025, 53),4)
          print ("Valor na t de student: ", t)
```

Valor na t de student: 2.0057

```
In [367]: # Valor limite inferior
          vli = 166 - t * math.sqrt(variancia)/math.sqrt(54)
          vli
```

Out[367]: 163.16416852843793

```
In [368]: # Valor limite superior
          vls = 166 + t * math.sqrt(variancia)/math.sqrt(54)
          vls
```

Out[368]: 168.83583147156207

A média amostral é 167.44 e portanto não está na região de rejeição e a hipótese nula não é rejeitada ao nível de 5% de significância.

A média populacional da altura é 168.15, maior que 166 cm, ou seja, houve um erro do tipo II.

c) $H_0 : \mu_F = \mu_M$ vs $H_1 : \mu_F \neq \mu_M$, com nível de significância igual a 1%, para a variável IMC.

```
In [369]: # Para mulheres
          media = df.IMC_am.loc[(df.Sexo_am == 'F')].mean()
          desvio_padrao = df.IMC_am.loc[(df.Sexo_am == 'F')].std()
          variancia = round(std.variance(df.IMC_am.loc[(df.Sexo_am == 'F')]),2)
          amostra = df.loc[(df.Sexo_am == 'F')].shape[0]

          print("Média do IMC das mulheres: ", round(media,2), "\nDesvio Padrão: ",
                round(desvio_padrao,2), "\nVariância: ", variancia, "\nAmostra: ", amostr
```

Média do IMC das mulheres: 22.48
Desvio Padrão: 5.03
Variância: 25.3
Amostra: 24

```
In [370]: # Para homens
media_h = df.IMC_am.loc[(df.Sexo_am == 'M')].mean()
desvio_padrao_h = df.IMC_am.loc[(df.Sexo_am == 'M')].std()
variancia_h = round(std.variance(df.IMC_am.loc[(df.Sexo_am == 'M')]),2)
amostra_h = df.loc[(df.Sexo_am == 'M')].shape[0]

print("Média do IMC dos homens: ", round(media_h,2), "\nDesvio Padrão: ",
      round(desvio_padrao_h,2), "\nVariância: ", variancia_h, "\nAmostra: ", am
```

```
Média do IMC dos homens:  25.14
Desvio Padrão:  4.7
Variância:  22.06
Amostra:  30
```

```
In [371]: # obtenção da variância conjunta
n_A = 24
n_B = 30
s_A = 25.3
s_B = 22.06
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B)/(n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

```
S_p =  4.846965743955379
```

```
In [372]: valor = s_p * math.sqrt((1/24) + (1/30))
print(round(valor,2))
```

```
1.33
```

```
In [373]: # t de student para 1% de significância
#g.l = 54 - 2 = 52
t = round(stats.t.ppf(1-0.005, 52),4)
print ("Valor na t de student: ", t)
```

```
Valor na t de student:  2.6737
```

```
In [374]: t * valor
```

```
Out[374]: 3.5490593180904177
```

RA = [-3.54 ; 3.54]

```
In [375]: # diferença das médias amostrais
dif_media = media - media_h
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: -2.66

Podemos observar que a média amostral está na região de aceitação, e a hipótese nula não é rejeitada.

Observando os dados populacionais, a média amostral de IMC em mulheres é aproximadamente 22.6 e em homens 23.8.

d) $H_0 : \mu M = \mu F$ vs $H_1 : \mu M > \mu F$, com nível de significância igual a 5%, para a variável PESO.

```
In [376]: # Para homens
media_h = df.Peso_am.loc[(df.Sexo_am == 'M')].mean()
desvio_padrao_h = df.Peso_am.loc[(df.Sexo_am == 'M')].std()
variância_h = round(std.variance(df.Peso_am.loc[(df.Sexo_am == 'M')]),2)
amostra_h = df.loc[(df.Sexo_am == 'M')].shape[0]

print("Média de Peso dos homens: ", round(media_h,2), "\nDesvio Padrão: ",
      round(desvio_padrao_h,2), "\nVariância: ", variância_h, "\nAmostra: ", am
```

Média de Peso dos homens: 73.27
Desvio Padrão: 15.39
Variância: 236.75
Amostra: 30

```
In [377]: # Para mulheres
media = df.Peso_am.loc[(df.Sexo_am == 'F')].mean()
desvio_padrao = df.Peso_am.loc[(df.Sexo_am == 'F')].std()
variancia = round(std.variance(df.Peso_am.loc[(df.Sexo_am == 'F')]),2)
amostra = df.loc[(df.Sexo_am == 'F')].shape[0]

print("Média de Peso das mulheres: ", round(media,2), "\nDesvio Padrão: ",
      round(desvio_padrao,2),"\nVariância: ",variancia, "\nAmostra: ", amostr
```

```
Média de Peso das mulheres:  60.5
Desvio Padrão:  15.41
Variância:  237.48
Amostra:  24
```

```
In [378]: # obtenção da variância conjunta
n_A = 30
n_B = 24
s_A = 236.75
s_B = 237.48
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B)/(n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

```
S_p =  15.397171318634623
```

```
In [379]: valor = s_p * math.sqrt((1/30) + (1/24))
print(round(valor,2))
```

```
4.22
```

```
In [380]: # t de student para 5% de significância
#g.l = 54 - 2 = 52
t = round(stats.t.ppf(1-0.025, 52),4)
print ("Valor na t de student: ", t)
```

```
Valor na t de student:  2.0066
```

```
In [381]: t * valor
```

```
Out[381]: 8.461208200562606
```

```
RA = [-8.46 ; 8.46]
```

```
In [382]: # diferença das médias amostrais
dif_media = media_h - media
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: 12.77

Podemos observar que a média amostral está na região de rejeição e a hipótese nula é rejeitada. A hipótese alternativa no entanto, não é rejeitada.

Se observamos os dados populacionais, a média de peso para homens é aproximadamente 69.7, enquanto para mulheres 62.1. Logo, de fato, a média dos homens é maior.

e) $H_0 : \mu_{UFAC} = \mu_{UFSCAR}$ vs $H_1 : \mu_{UFAC} \neq \mu_{UFSCAR}$, com nível de significância igual a 10%, para a variável IDADE.

```
In [383]: # Para UFAC
media = df.Idade_am.loc[(df.IES_am == 'UFAC')].mean()
desvio_padrao = df.Idade_am.loc[(df.IES_am == 'UFAC')].std()
variancia = round(std.variance(df.Idade_am.loc[(df.IES_am == 'UFAC')]),2)
amostra = df.loc[(df.IES_am == 'UFAC')].shape[0]

print("Média de idade dos alunos da UFAC ", round(media,2), "\nDesvio Padrão:
      round(desvio_padrao,2),"\nVariância: ",variancia, "\nAmostra: ", amostr
```

Média de idade dos alunos da UFAC 20.9
Desvio Padrão: 2.51
Variância: 6.31
Amostra: 20


```
In [384]: # Para UFSCAR
media_uf = df.Idade_am.loc[(df.IES_am == 'UFSCAR')].mean()
desvio_padrao_uf = df.Idade_am.loc[(df.IES_am == 'UFSCAR')].std()
variancia_uf = round(std.variance(df.Idade_am.loc[(df.IES_am == 'UFSCAR')]),2)
amostra_uf = df.loc[(df.IES_am == 'UFSCAR')].shape[0]
```

```
print("Média de idade dos alunos da UFSCAR ", round(media_uf,2), "\nDesvio Pa
      round(desvio_padrao_uf,2),"\nVariância: ",variancia_uf, "\nAmostra: ",
```

```
Média de idade dos alunos da UFSCAR  22.21
Desvio Padrão:  4.18
Variância:  17.5
Amostra:  34
```

```
In [385]: # obtenção da variância conjunta
n_A = 20
n_B = 34
s_A = 6.31
s_B = 17.5
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B)/(n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

```
S_p =  3.6621504821410813
```

```
In [386]: valor = s_p * math.sqrt((1/20) + (1/34))
print(round(valor,2))
```

```
1.03
```

```
In [387]: # t de student para 10% de significância
#g.l = 54 - 2 = 52
t = round(stats.t.ppf(1-0.05, 52),4)
print ("Valor na t de student: ", t)
```

```
Valor na t de student:  1.6747
```

```
In [388]: t * valor
```

```
Out[388]: 1.728286071380589
```

```
RA = [-1.72 ; 1.72]
```

```
In [389]: # diferença das médias amostrais
dif_media = media - media_uf
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: -1.31

A diferença das médias amostrais está dentro da RA, portanto a hipótese nula não é rejeitada. Nos dados populacionais, a média da idade dos alunos da UFAC é 20.9, enquanto da ufscar é 22.

Podemos perceber que 10% de erro é um valor significativo.

PARTE B)

Utilizando a amostra estratificada, por SEXO, verificar se existem diferenças significativas, ao nível de 5% de significância (PARA CADA SEXO – M ou F):

a) Entre as alturas dos alunos da UFSCAR e da UFAC.

Mulheres

```
In [390]: # importar amostra estratificada por sexo e exibir as 5 primeiras entradas
df = pd.read_csv("amostra_sexo_ok.csv", sep = ';')
df.IMC = df.IMC.apply(lambda x: x.replace(',', '.'))
df.IMC = df.IMC.astype(float)
df.head()
```

Out[390]:

	id	peso	altura	IES	sexo	IMC
0	19	78	181	UFAC	F	25.55358
1	19	68	159	UFAC	F	17.19108
2	19	59	177	UFAC	F	18.48411
3	20	72	163	UFSCAR	F	19.12968
4	20	63	168	UFAC	F	17.78112

```
In [391]: # verificar a proporção de homens e mulheres
df.sexo.value_counts()/df.shape[0]
```

```
Out[391]: M    0.518519
          F    0.481481
          Name: sexo, dtype: float64
```

Parâmetro de interesse a ser testado: Se μ_F UFSCAR = μ_F UFAC, conseqüentemente $\mu_A - \mu_B = 0$
Hipótese nula: $H_0: \mu_A - \mu_B = 0$
Hipótese alternativa: $H_1: \mu_A - \mu_B \neq 0$
Estatística do teste:

```
In [392]: # Para mulheres alunas da UFSCAR
media = df.altura.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].mean()
desvio_padrao = df.altura.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].std()
variancia = round(std.variance(df.altura.loc[(df.sexo == 'F') & (df.IES == 'U
amostra = df.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].shape[0]

print("Média da altura das alunas da UFSCAR: ", round(media,2), "\nDesvio Pad
round(desvio_padrao,2),"\nVariância: ",variancia, "\nAmostra: ", amostr
```

Média da altura das alunas da UFSCAR: 167.72
Desvio Padrão: 7.51
Variância: 56.33
Amostra: 18

```
In [393]: # Para mulheres alunas da UFAC
media_ufac = df.altura.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].mean()
desvio_padrao_ufac = df.altura.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].std
variancia_ufac = round(std.variance(df.altura.loc[(df.sexo == 'F') & (df.IES
amostra_ufac = df.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].shape[0]

print("Média da altura das alunas da UFAC: ", round(media_ufac,2), "\nDesvio
round(desvio_padrao_ufac,2),"\nVariância: ",variancia_ufac, "\nAmostra:
```

Média da altura das alunas da UFAC: 168.25
Desvio Padrão: 9.51
Variância: 90.5
Amostra: 8

Cálculos:

```
In [394]: # obtenção da variância conjunta
n_A = 18
n_B = 8
s_A = 56.33
s_B = 90.5
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B)/(n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

S_p = 8.142250917283254

O grau de liberdade será 24, pois $18 + 8 - 2 = 24$

```
In [395]: # t de student para 5% de significância
t = round(stats.t.ppf(1-0.025, 24),4)
print ("Valor na t de student: ", t)
```

Valor na t de student: 2.0639

```
In [396]: valor = s_p * math.sqrt((1/18) + (1/8))
print(round(valor,2))
```

3.46

RR = (-infinito; - 2.0639 3.46) U (2.0639 3.46; + infinito)

```
In [397]: t * valor
```

Out[397]: 7.140663386492991

RR = (- infinito ; -7.14) U (7.14 ; + infinito)

RA = [-7.14; 7.14]

```
In [398]: # diferença das médias amostrais
dif_media = media - media_ufac
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: -0.53

A diferença das médias amostrais pertence à RA, portanto não rejeitamos a hipótese.

Homens

```
In [399]: # Para homens alunos da UFSCAR
media = df.altura.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].mean()
desvio_padrao = df.altura.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].std()
variancia = round(std.variance(df.altura.loc[(df.sexo == 'M') & (df.IES == 'U
amostra = df.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].shape[0]

print("Média da altura dos alunos da UFSCAR: ", round(media,2), "\nDesvio Pad
round(desvio_padrao,2),"\nVariância: ",variancia, "\nAmostra: ", amostr
```

Média da altura dos alunos da UFSCAR: 172.11
Desvio Padrão: 9.15
Variância: 83.63
Amostra: 18

```
In [400]: # Para homens alunos da UFAC
media_ufac = df.altura.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].mean()
desvio_padrao_ufac = df.altura.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].std
variancia_ufac = round(std.variance(df.altura.loc[(df.sexo == 'M') & (df.IES
amostra_ufac = df.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].shape[0]

print("Média da altura dos alunos da UFAC: ", round(media_ufac,2), "\nDesvio
round(desvio_padrao_ufac,2),"\nVariância: ",variancia_ufac, "\nAmostra:
```

Média da altura dos alunos da UFAC: 161.1
Desvio Padrão: 4.36
Variância: 18.99
Amostra: 10

Cálculos:

```
In [401]: # obtenção da variância conjunta
n_A = 18
n_B = 10
s_A = 83.63
s_B = 18.99
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B)/(n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

S_p = 7.826532781801619

O grau de liberdade será 26, pois $18 + 10 - 2 = 26$

```
In [402]: # t de student para 5% de significância
t = round(stats.t.ppf(1-0.025, 26),4)
print ("Valor na t de student: ", t)
```

Valor na t de student: 2.0555

```
In [403]: valor = s_p * math.sqrt((1/18) + (1/10))
print(round(valor,2))
```

3.09

RR = (-infinito; - 2.0555 3.09) U (2.0555 3.09; + infinito)

```
In [404]: t * valor
```

Out[404]: 6.344971166697805

RR = (- infinito ; -6.345) U (6.345 ; + infinito)

RA = [-6.345 ; 6.345]

```
In [405]: # diferença das médias amostrais
dif_media = media - media_ufac
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: 11.01

A diferença das médias pertence à região de rejeição, portanto, rejeitamos a hipótese nula.

Observando os dados populacionais, a média da altura da UFSCAR é 171 e UFAC 162, portanto há uma diferença considerável. Todavia, as maiores discrepâncias ocorrem nos homens, se observarmos o boxplot.

b) Entre os IMCs dos alunos da UFSCAR e da UFAC.

Mulheres

```
In [406]: # Para mulheres com alunos da UFSCAR
media = df.IMC.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].mean()
desvio_padrao = df.IMC.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].std()
variancia = round(std.variance(df.IMC.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')], 2)
amostra = df.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].shape[0]
```

```
print("Média do IMC das alunas da UFSCAR: ", round(media,2), "\nDesvio Padrão: ",
      round(desvio_padrao,2), "\nVariância: ", variancia, "\nAmostra: ", amostra)
```

```
Média do IMC das alunas da UFSCAR:  20.38
Desvio Padrão:  4.42
Variância:  19.56
Amostra:  18
```

```
In [407]: # Para mulheres com alunos da UFAC
media_ufac = df.IMC.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].mean()
desvio_padrao_ufac = df.IMC.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].std()
variancia_ufac = round(std.variance(df.IMC.loc[(df.sexo == 'F') & (df.IES == 'UFAC')], 2)
amostra_ufac = df.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].shape[0]
```

```
print("Média da IMC das alunas da UFAC: ", round(media_ufac,2), "\nDesvio Padrão: ",
      round(desvio_padrao_ufac,2), "\nVariância: ", variancia_ufac, "\nAmostra: ", amostra_ufac)
```

```
Média da IMC das alunas da UFAC:  18.26
Desvio Padrão:  4.42
Variância:  19.51
Amostra:  8
```

```
In [408]: # obtenção da variância conjunta
n_A = 18
n_B = 8
s_A = 19.56
s_B = 19.51
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B)/(n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

S_p = 4.421019867255367

O grau de liberdade será 24, pois $18 + 8 - 2 = 24$

```
In [409]: # t de student para 5% de significância
t = round(stats.t.ppf(1-0.025, 24),4)
print ("Valor na t de student: ", t)
```

Valor na t de student: 2.0639

```
In [410]: valor = s_p * math.sqrt((1/18) + (1/8))
print(round(valor,2))
```

1.88

RR = (-infinito; - 2.0639 1.88) U (2.0639 1.88; + infinito)

```
In [411]: t * valor
```

Out[411]: 3.8771851933545958

RR = (-infinito ; -3.88) U (3.88 ; + infinito)

RA = [-3.88 ; 3.88]

```
In [412]: # diferença das médias amostrais
dif_media = media - media_ufac
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: 2.11

Como a diferença das médias amostrais está dentro do RA, não rejeitamos a hipótese nula.

Homens

```
In [413]: # Para homens alunos da UFSCAR
media = df.IMC.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].mean()
desvio_padrao = df.IMC.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].std()
variancia = round(std.variance(df.IMC.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')], 2)
amostra = df.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].shape[0]

print("Média da IMC dos alunos da UFSCAR: ", round(media,2), "\nDesvio Padrão: ",
      round(desvio_padrao,2), "\nVariância: ", variancia, "\nAmostra: ", amostra)
```

Média da IMC dos alunos da UFSCAR: 21.42
Desvio Padrão: 5.1
Variância: 26.0
Amostra: 18

```
In [414]: # Para homens alunos da UFAC
media_ufac = df.IMC.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].mean()
desvio_padrao_ufac = df.IMC.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].std()
variancia_ufac = round(std.variance(df.IMC.loc[(df.sexo == 'M') & (df.IES == 'UFAC')], 2)
amostra_ufac = df.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].shape[0]

print("Média da IMC dos alunos da UFAC: ", round(media_ufac,2), "\nDesvio Padrão: ",
      round(desvio_padrao_ufac,2), "\nVariância: ", variancia_ufac, "\nAmostra: ", amostra_ufac)
```

Média da IMC dos alunos da UFAC: 14.92
Desvio Padrão: 4.34
Variância: 18.82
Amostra: 10

```
In [415]: # obtenção da variância conjunta
n_A = 18
n_B = 10
s_A = 26.0
s_B = 18.82
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B)/(n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

S_p = 4.849187084926234

O grau de liberdade será 24, pois $18 + 10 - 2 = 26$

```
In [416]: # t de student para 5% de significância
t = round(stats.t.ppf(1-0.025, 26),4)
print ("Valor na t de student: ", t)
```

Valor na t de student: 2.0555

```
In [417]: valor = s_p * math.sqrt((1/18) + (1/10))
print(round(valor,2))
```

1.91

RR = (-infinito; - 2.0555 1.91) U (2.0555 1.91; + infinito)

```
In [418]: t * valor
```

Out[418]: 3.931236614420433

RR = (-infinito ; -3.93) U (3.93 ; + infinito)

RA = [-3.93 ; 3.93]

```
In [419]: # diferença das médias amostrais
dif_media = media - media_ufac
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: 6.5

Como a diferença das médias amostrais não pertence a RA, rejeitamos a hipótese nula.

PARTE C: Utilizando a amostra estratificada, por IES, verificar se existem diferenças significativas, ao nível de 5% de significância (DENTRO DE CADA IES):

A) Entre os pesos dos alunos M e F.

```
In [420]: # importar amostra estratificada por sexo e exibir as 5 primeiras entradas
df = pd.read_csv("amostra_IES_ok.csv", sep = ';')
df.head()
```

Out[420]:

	sexo	idade	peso	altura	IES
0	M	20	51	164	UFAC
1	M	26	59	156	UFAC
2	M	15	69	158	UFAC
3	M	22	42	152	UFAC
4	F	19	59	177	UFAC

```
In [421]: # verificar a proporção de IES
df.IES.value_counts()/df.shape[0]
```

```
Out[421]: UFSCAR    0.611111
UFAC          0.388889
Name: IES, dtype: float64
```

Parâmetro de interesse a ser testado: Se $\mu_M = \mu_F$, consequentemente $\mu_A - \mu_B = 0$

Hipótese nula: $H_0: \mu_A - \mu_B = 0$

Hipótese alternativa: $H_1: \mu_A - \mu_B \neq 0$

Estatística do teste:

UFSCAR

```
In [422]: # Para homens alunos da UFSCAR
media = df.peso.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].mean()
desvio_padrao = df.peso.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].std()
variancia = round(std.variance(df.peso.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')], 2)
amostra = df.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].shape[0]

print("Média de peso dos alunos da UFSCAR: ", round(media,2), "\nDesvio Padrão: ", round(desvio_padrao,2), "\nVariância: ", variancia, "\nAmostra: ", amostra)
```

Média de peso dos alunos da UFSCAR: 74.72
Desvio Padrão: 13.45
Variância: 180.8
Amostra: 18

```
In [423]: # Para alunas da UFSCAR
media_f = df.peso.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].mean()
desvio_padrao_f = df.peso.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].std()
variancia_f = round(std.variance(df.peso.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')], 2)
amostra_f = df.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].shape[0]

print("Média de Peso das alunas da UFSCAR: ", round(media_f,2), "\nDesvio Padrão: ", round(desvio_padrao_f,2), "\nVariância: ", variancia_f, "\nAmostra: ", amostra_f)
```

Média de Peso das alunas da UFSCAR: 60.73
Desvio Padrão: 9.6
Variância: 92.21
Amostra: 15

```
In [424]: # obtenção da variância conjunta
n_A = 18
n_B = 15
s_A = 180.8
s_B = 92.21
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B) / (n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

S_p = 11.865564162871726

O grau de liberdade será 31, pois $18 + 15 - 2 = 31$

```
In [425]: # t de student para 5% de significância
t = round(stats.t.ppf(1-0.025, 31),4)
print ("Valor na t de student: ", t)
```

Valor na t de student: 2.0395

```
In [426]: valor = s_p * math.sqrt((1/18) + (1/15))
print(round(valor,2))
```

4.15

RR = (-infinito; - 2.0395 4.15) U (2.0395 4.15; + infinito)

```
In [427]: t * valor
```

Out[427]: 8.460327786020596

RR = (-infinito ; -8.46) U (8.46 ; + infinito)

RA = [-8.46 ; 8.46]

```
In [428]: # diferença das médias amostrais
dif_media = media - media_f
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: 13.99

Como 13.99 não pertence à RA, então rejeitamos a hipótese nula.

UFAC

```
In [429]: # Para homens alunos da UFAC
media = df.peso.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].mean()
desvio_padrao = df.peso.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].std()
variancia = round(std.variance(df.peso.loc[(df.sexo == 'M') & (df.IES == 'UFA
amostra = df.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].shape[0]

print("Média de peso dos alunos da UFAC: ", round(media,2), "\nDesvio Padrão:
round(desvio_padrao,2),"\nVariância: ",variancia, "\nAmostra: ", amostr
```

Média de peso dos alunos da UFAC: 56.27
Desvio Padrão: 9.01
Variância: 81.22
Amostra: 11

```
In [430]: # Para alunas da UFAC
media_f = df.peso.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].mean()
desvio_padrao_f = df.peso.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].std()
variancia_f = round(std.variance(df.peso.loc[(df.sexo == 'F') & (df.IES == 'U
amostra_f = df.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].shape[0]

print("Média de Peso das alunas da UFAC: ", round(media_f,2), "\nDesvio Padrã
round(desvio_padrao_f,2),"\nVariância: ",variancia_f, "\nAmostra: ", am
```

Média de Peso das alunas da UFAC: 56.9
Desvio Padrão: 7.36
Variância: 54.1
Amostra: 10

```
In [431]: # obtenção da variância conjunta
n_A = 11
n_B = 10
s_A = 81.22
s_B = 54.1
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B)/(n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

S_p = 8.268838141512163

O grau de liberdade será 19, pois $10 + 11 - 2 = 19$

```
In [432]: # t de student para 5% de significância
t = round(stats.t.ppf(1-0.025, 19),4)
print ("Valor na t de student: ", t)
```

Valor na t de student: 2.093

```
In [433]: valor = s_p * math.sqrt((1/11) + (1/10))
print(round(valor,2))
```

3.61

RR = (-infinito; - 2.093 3.61) U (2.093 3.61; + infinito)

```
In [434]: t * valor
```

Out[434]: 7.561831992210688

RR = (-infinito ; --7.56) U (7.56 ; + infinito)

RA = [-7.56 ; 7.56]

```
In [435]: # diferença das médias amostrais
dif_media = media - media_f
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: -0.63

Como a média amostral pertence à RA, não rejeitamos a hipótese nula.

b) Entre as idades dos alunos M e F.

UFSCAR

```
In [436]: # Para homens alunos da UFSCAR
media = df.idade.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].mean()
desvio_padrao = df.idade.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')].std()
variancia = round(std.variance(df.idade.loc[(df.sexo == 'M') & (df.IES == 'UF
amostra = df.loc[(df.sexo == 'M') & (df.IES == 'UFSCAR')]).shape[0]
```

```
print("Média de idade dos alunos da UFSCAR: ", round(media,2), "\nDesvio Padr
round(desvio_padrao,2),"\nVariância: ",variancia, "\nAmostra: ", amostr
```

```
Média de idade dos alunos da UFSCAR:  21.39
Desvio Padrão:  3.52
Variância:  12.37
Amostra:  18
```

```
In [437]: # Para alunas da UFSCAR
media_f = df.idade.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].mean()
desvio_padrao_f = df.idade.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')].std()
variancia_f = round(std.variance(df.idade.loc[(df.sexo == 'F') & (df.IES == '
amostra_f = df.loc[(df.sexo == 'F') & (df.IES == 'UFSCAR')]).shape[0]
```

```
print("Média de idade das alunas da UFSCAR: ", round(media_f,2), "\nDesvio Pa
round(desvio_padrao_f,2),"\nVariância: ",variancia_f, "\nAmostra: ", am
```

```
Média de idade das alunas da UFSCAR:  21.8
Desvio Padrão:  3.19
Variância:  10.17
Amostra:  15
```

```
In [438]: # obtenção da variância conjunta
n_A = 18
n_B = 15
s_A = 12.37
s_B = 10.17
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B)/(n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

```
S_p =  3.372899585357267
```

O grau de liberdade será 31, pois $18 + 15 - 2 = 31$


```
In [439]: # t de student para 5% de significância
t = round(stats.t.ppf(1-0.025, 31),4)
print ("Valor na t de student: ", t)
```

Valor na t de student: 2.0395

```
In [440]: valor = s_p * math.sqrt((1/18) + (1/15))
print(round(valor,2))
```

1.18

RR = (-infinito ; -2.0395 1.18) U (2.0395 1.18; + infinito)

```
In [441]: t * valor
```

Out[441]: 2.4049287239747343

RR = (-infinito ; -2.40) U (2.40 ; + infinito)

RA = [-2.40 ; 2.40]

```
In [442]: # diferença das médias amostrais
dif_media = media - media_f
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: -0.41

Como a diferença das médias amostrais pertence à RA, não rejeitamos a hipótese nula.

UFAC

```
In [443]: # Para homens alunos da UFAC
media = df.idade.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].mean()
desvio_padrao = df.idade.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].std()
variancia = round(std.variance(df.idade.loc[(df.sexo == 'M') & (df.IES == 'UFAC')], 2)
amostra = df.loc[(df.sexo == 'M') & (df.IES == 'UFAC')].shape[0]

print("Média de idade dos alunos da UFAC: ", round(media,2), "\nDesvio Padrão: ", round(desvio_padrao,2), "\nVariância: ", variancia, "\nAmostra: ", amostra)
```

Média de idade dos alunos da UFAC: 20.64
Desvio Padrão: 3.38
Variância: 11.45
Amostra: 11

```
In [444]: # Para alunas da UFAC
media_f = df.idade.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].mean()
desvio_padrao_f = df.idade.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].std()
variancia_f = round(std.variance(df.idade.loc[(df.sexo == 'F') & (df.IES == 'UFAC')], 2)
amostra_f = df.loc[(df.sexo == 'F') & (df.IES == 'UFAC')].shape[0]

print("Média de idade das alunas da UFAC: ", round(media_f,2), "\nDesvio Padrão: ", round(desvio_padrao_f,2), "\nVariância: ", variancia_f, "\nAmostra: ", amostra_f)
```

Média de idade das alunas da UFAC: 20.0
Desvio Padrão: 1.89
Variância: 3.56
Amostra: 10

```
In [445]: # obtenção da variância conjunta
n_A = 11
n_B = 10
s_A = 11.45
s_B = 3.56
s2_p = ((n_A - 1) * s_A + (n_B - 1) * s_B) / (n_A + n_B - 2)
s_p = math.sqrt(s2_p)
print("S_p = ", s_p)
```

S_p = 2.7771625049584996

O grau de liberdade será 19, pois $10 + 11 - 2 = 19$

```
In [446]: # t de student para 5% de significância
t = round(stats.t.ppf(1-0.025, 19),4)
print ("Valor na t de student: ", t)
```

Valor na t de student: 2.093

```
In [447]: valor = s_p * math.sqrt((1/11) + (1/10))
print(round(valor,2))
```

1.21

RR = (-infinito; - 2.093 1.21) U (2.093 1,21; + infinito)

```
In [448]: t * valor
```

Out[448]: 2.5397082296404343

RR = (-infinito ; -2.53) U (2.53 ; + infinito)

RA = [-2.53 ; 2.53]

```
In [449]: # diferença das médias amostrais
dif_media = media - media_f
print("Diferença das médias amostrais: ", round(dif_media,2))
```

Diferença das médias amostrais: 0.64

Como a diferença das médias pertence à RA, não podemos rejeitar a hipótese nula.

Dados os resultados obtidos, não podemos concluir que uma amostra como um todo é melhor que a outra. A amostra estratificada por sexo, por exemplo, traz dados mais coerentes quando observamos peso e altura pois diferenciar o sexo em grupos é um dado relevante. Contudo, podemos observar que os resultados estão coerentes num todo à realidade da população.