



# Monitoria Trabalho COVID 1



# Instalação do Microsoft Visual Studio e Python 3.9.0

- Vídeo tutorial do Khennedy;
- <https://pandas.pydata.org/>
- <https://visualstudio.microsoft.com/pt-br/> (Professional)
- <https://www.python.org/>
- Instalação no Linux

# Arquivos .csv

- Podem ser abertos em diferentes programas, como Microsoft Excell ou LibreOffice;
- Delimitador dos arquivos .csv: “;”
- Limpar dados inconscientes e manter um arquivo para registrar as modificações

dados-ac.csv - LibreOffice Calc

Arquivo Editar Exibir Inserir Formatar Estilos Planilha Dados Ferramentas Janela Ajuda

Liberation Sans 10 pt N I S A % 00 00 00

C6  $f_x \Sigma =$  2020-03-18T05:00:00.000Z

	A	B	C	D	
1	jid	dataNotificacao	dataInicioSintomas	dataNascimento	sintomas
2	3ODKsTM6lq	2020-04-20T03:00:00.000Z	2020-04-15T03:00:00.000Z	undefined	Dor de Garganta, Tosse
3	lqffBRttV	2020-04-21T17:00:15.391Z	2020-04-13T05:00:00.000Z	undefined	Outros
4	pi0EZEfQCF	2020-04-22T09:12:45.446Z	2020-04-12T04:00:00.000Z	undefined	Febre, Tosse, Outros
5	t6jUD1xWg2	2020-03-18T05:00:00.000Z	2020-03-17T05:00:00.000Z	undefined	Febre, Tosse, Outros
6	EYngeYglcP	2020-03-18T05:00:00.000Z	2020-03-18T05:00:00.000Z	undefined	Tosse, Febre, Dor de Garganta, Dispneia, Outros



# Código Python

```
import pandas as pd
df = pd.read_csv ('dados-ac-modificados.csv', delimiter=";")

# medidas
mean1 = df['idade'].mean()
sum1 = df['idade'].sum()
max1 = df['idade'].max()
min1 = df['idade'].min()
count1 = df['idade'].count()
median1 = df['idade'].median()
std1 = df['idade'].std()
var1 = df['idade'].var()
```



# Código Python

```
# print das medidas
print ('Media das idades: {}'.format(mean1))
print ('Soma das idades: {}'.format(sum1))
print ('Idade maxima: {}'.format(max1))
print ('Idade minima: {}'.format(min1))
print ('Coontagem das idades: {}'.format(count1))
print ('Mediana das idades: {}'.format(median1))
print ('Desvio padrao de idade: {}'.format(std1))
print ('Variancia de: {}'.format(var1))

# contagem

count = df['evolucaoCaso'].value_counts()
print('Evolucao dos casos = \n{}'.format(count))
```



## Exemplo 1: Cálculo da frequência relativa a partir da saída

Através dos dados obtidos na saída do nosso programa, podemos calcular, por exemplo, a frequência relativa de cura no estado selecionado (Acre):

$$\text{Frequência relativa} = \frac{\text{Frequência absoluta}}{\text{N}^\circ \text{ total de observações}}$$

$$\text{Frequência relativa} = \frac{37548}{223705} \cong 0.1678 = 16,78\%$$

Se tivéssemos infinitas repetições do experimento “Evolução do caso” (ou seja,  $N^\circ$  total de observações  $\rightarrow \infty$ ) e um conjunto  $\Omega$  de elementos equiprováveis e mutuamente exclusivos, poderíamos calcular a probabilidade do evento “Cura” de acontecer.



## Exemplo 2: Probabilidade Condicional

Tratando nossos valores de frequências relativas como se fossem estimativas das probabilidades (que só poderiam ser obtidas com infinitas repetições do experimento) baseados na amostra, vamos verificar a relação dos eventos  $A = \text{Cura}$  e  $B = \text{Paciente apresentou 2 sintomas}$ .

$$P(B = \text{Apresentou 2 sintomas}) = \frac{152109}{223705} \cong 0,6799$$

Aproximando a frequência relativa para uma estimativa da probabilidade de uma pessoa ter sido curada e ter apresentado 2 sintomas é:

$$P(A \cap B) = 0.042962$$

Assim, podemos obter um valor para a probabilidade condicional que nos diz qual a probabilidade do paciente que apresentou 2 sintomas ser curado:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.042962}{0.6799} \cong 0,06318 = 6,318\%$$

Se os dados fossem de toda a população, poderíamos considerar a estimativa obtida como o valor da probabilidade.



## Grupos, arquivos .csv e próxima monitoria

- Organizar a turma em grupos de 3 a 4 pessoas;
- Cada grupo escolher um estado;
- Repetir o código apresentado para seu próprio estado em R ou em Python até a próxima monitoria (dia 17);
- Dúvidas: [rvbrant@usp.br](mailto:rvbrant@usp.br) e [khennedy@usp.br](mailto:khennedy@usp.br)



## Referências

1. Aulas de “SME0620 - Estatística I” da Profa. Dra. Mariana Curi
2. <https://datatofish.com/use-pandas-to-calculate-stats-from-an-imported-csv-file/>
3. <https://opendatasus.saude.gov.br/dataset/casos-nacionais>