

Ciência de Dados

Licenciatura Engenharia Informática
2º Semestre – 2021/2022

Ricardo Jesus Ferreira
ricardojesus.ferreira@my.istec.pt

Aula Prática

Numpy

Pandas

Seaborn

Matplotlib

Scipy

Dataset - Iris

- 150 linhas
- Download a partir da UCI Machine Learning Repository
- Atributos da informação
 - Caule
 - Comprimento e largura
 - Pétalas
 - Comprimento e largura
 - Espécie

SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3.0	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5.0	3.6	1.4	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
5.0	3.4	1.5	0.2	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa
5.4	3.7	1.5	0.2	Iris-setosa
4.8	3.4	1.6	0.2	Iris-setosa
4.8	3.0	1.4	0.1	Iris-setosa
4.3	3.0	1.1	0.1	Iris-setosa

Análise Estatística

- Apresente a média das 4 características presentes no *dataset*
- Apresente a mediana das 4 características presentes no *dataset*
- Apresente a moda das 4 características presentes no *dataset*
- Apresente a amplitude das 4 características presentes no *dataset*
- Apresente a variância das 4 características presentes no *dataset*

Análise Estatística

- Apresente num gráfico a distribuição das observações das 4 características presentes no *dataset*
- Apresente individualmente a distribuição das observações de cada característica presentes no *dataset*
- Assumindo um intervalo de confiança de 90%, quais são os limites superior e inferior de cada característica

Bibliografia

- B. Gomez,(2020) “Resolviendo problemas de Big Data”, Alfaomega.
- D. Insua, (2019)“Big data: Conceptos, tecnologías y aplicaciones”, CSIC.
- H. Jones, (2019)“Analítica de datos”, HJ,.
- J. Somed, (2020)“Big Data Analytics”, JLC.
- D. Petković (2020)“Microsoft® SQL Server® 2019 A Beginner’s Guide - Seventh Edition”, McGraw Hill.