AULA 11 - EXERCÍCIO PRÁTICO - T2 - CPA

Leonardo Miranda e Pedro Barcelos

I) Etapa Teórica

1) **Quais atributos são categóricos e quais atributos são numéricos? Apenas com as informações dadas é possível identificar dentre os atributos numéricos quais são os tipos?**

Os atributos categóricos estão nas colunas **f2** e **f3**, enquanto os numéricos nas colunas **f1** e **f4**. É impossível identificar exatamente os tipos dos atributos numéricos pois precisamos de um contexto para realizar essa identificação.

---

2) **Qual a moda de f3:**

Tendo em vista que a moda é o valor mais frequente em um conjunto, podemos afirmar que a moda de f3 é ‘red’, pois aparece 6 vezes, enquanto ‘black’ e ‘yellow’ aparecem apenas 2 vezes cada um.

---

3) **Compute a média, mediana, desvio padrão, variância, dos atributos f1 e f4.**
4) **Calcule a correlação existente entre os atributos f1 e f4**

Para computar esses valores estatísticos, fizemos uma planilha no Excel. Segue abaixo as respostas das questões 3 e 4:

	A	B	C	D
1	**T2 - CPA**			
3	f1	f2	f3	f4
4	0,02	True	red	8,25
5	-2	False	red	1,35
6	0,57	True	yellow	21,86
7	1	False	black	3,06
8	-2	True	black	1,58
9	0,84	True	red	1,25
10	0,66	False	red	1,68
11	-1	False	yellow	1,01
12	0,44	True	red	1,19
13	0,32	False	red	1,47

H	I	J
Estatística		
	f1	f4
Média	-0,115	4,27
Mediana	0,38	1,525
D.P	1,078436368	6,21496259
Var	1,163025	38,62576
Correlação	0,251826137	

II) Etapa Prática

1) Considerando que a variável target refere-se a espécie do pinguim, quais são as features categóricas e numéricas que temos disponíveis?

Analisando os dados do arquivo, conclui-se as seguintes classificações:

Categóricas: studyName; Sample Number; Region; Island; Stage; Individual ID; Clutch Completion; Date Egg; Sex.

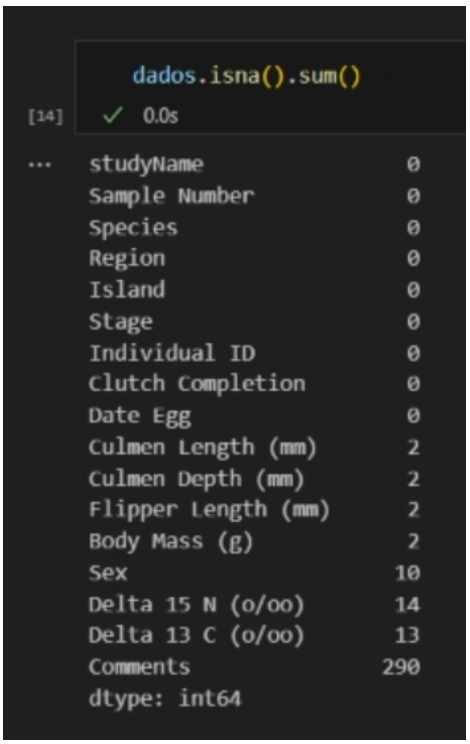
Numéricas: Culm Length (mm); Culmen Depth (mm); Flipper Length (mm); Body Mass (g); Delta 15 N (o/oo); Delta 13 C (o/oo).

Target: Species

---

2) Existem dados faltantes no conjunto de dados? Quantos e em quais features?

Podemos ver, através desses métodos da biblioteca Pandas, que existem sim dados faltantes no conjunto. Temos, abaixo, as colunas seguidas do número de dados faltantes nela. Por exemplo, na coluna ‘Body Mass (g)’ temos 2 dados faltando:



```
dados.isna().sum()
[14] ✓ 0.0s
...
studyName              0
Sample Number          0
Species                0
Region                 0
Island                 0
Stage                  0
Individual ID          0
Clutch Completion      0
Date Egg               0
Culmen Length (mm)     2
Culmen Depth (mm)      2
Flipper Length (mm)    2
Body Mass (g)          2
Sex                   10
Delta 15 N (o/oo)     14
Delta 13 C (o/oo)     13
Comments             290
dtype: int64
```

3) Quais espécies de pinguim existem no conjunto de dados? As classes estão balanceadas?

Vê-se, abaixo, que existem 3 espécies de pinguins no conjunto de dados (Adelie Penguin, Chinstrap penguin e Gentoo penguin). Pela disparidade do número de pinguins da classe Chinstrap em relação às outras, afirma-se que as classes não estão balanceadas: