190031920

A Nikhil Reddy

DS Practical 8

In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv('Iris.csv')
df.head()
```

Out[2]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

In [3]:
```
irisdf = df.loc[df["Species"] == "Iris-setosa"]
irisdf
```

Out[3]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa
10	11	5.4	3.7	1.5	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
13	14	4.3	3.0	1.1	0.1	Iris-setosa
14	15	5.8	4.0	1.2	0.2	Iris-setosa
15	16	5.7	4.4	1.5	0.4	Iris-setosa
16	17	5.4	3.9	1.3	0.4	Iris-setosa
17	18	5.1	3.5	1.4	0.3	Iris-setosa
18	19	5.7	3.8	1.7	0.3	Iris-setosa
19	20	5.1	3.8	1.5	0.3	Iris-setosa
20	21	5.4	3.4	1.7	0.2	Iris-setosa
21	22	5.1	3.7	1.5	0.4	Iris-setosa
22	23	4.6	3.6	1.0	0.2	Iris-setosa
23	24	5.1	3.3	1.7	0.5	Iris-setosa
24	25	4.8	3.4	1.9	0.2	Iris-setosa
25	26	5.0	3.0	1.6	0.2	Iris-setosa
26	27	5.0	3.4	1.6	0.4	Iris-setosa
27	28	5.2	3.5	1.5	0.2	Iris-setosa
28	29	5.2	3.4	1.4	0.2	Iris-setosa
29	30	4.7	3.2	1.6	0.2	Iris-setosa
30	31	4.8	3.1	1.6	0.2	Iris-setosa
31	32	5.4	3.4	1.5	0.4	Iris-setosa
32	33	5.2	4.1	1.5	0.1	Iris-setosa
33	34	5.5	4.2	1.4	0.2	Iris-setosa
34	35	4.9	3.1	1.5	0.1	Iris-setosa
35	36	5.0	3.2	1.2	0.2	Iris-setosa
36	37	5.5	3.5	1.3	0.2	Iris-setosa
37	38	4.9	3.1	1.5	0.1	Iris-setosa
38	39	4.4	3.0	1.3	0.2	Iris-setosa
39	40	5.1	3.4	1.5	0.2	Iris-setosa
40	41	5.0	3.5	1.3	0.3	Iris-setosa
41	42	4.5	2.3	1.3	0.3	Iris-setosa
42	43	4.4	3.2	1.3	0.2	Iris-setosa
43	44	5.0	3.5	1.6	0.6	Iris-setosa
44	45	5.1	3.8	1.9	0.4	Iris-setosa
45	46	4.8	3.0	1.4	0.3	Iris-setosa
46	47	5.1	3.8	1.6	0.2	Iris-setosa
47	48	4.6	3.2	1.4	0.2	Iris-setosa
48	49	5.3	3.7	1.5	0.2	Iris-setosa
49	50	5.0	3.3	1.4	0.2	Iris-setosa

In [4]:
```
print("Unique values in Sepal Length CM are :", irisdf['SepallengthCm'].unique())
```

```
Unique values in Sepal Length CM are : [5.1 4.9 4.7 4.6 5.  5.4 4.4 4.8 4.3 5.8 5.7 5.2 5.5 4.5 5.3]
```

In [5]:
```
sepallengthfrequency = irisdf.groupby('SepallengthCm').size()

sepallengthfrequency
```

```
SepalLengthCm
4.3    1
4.4    3
4.5    1
4.6    4
4.7    2
4.8    5
4.9    4
5.0    8
5.1    8
5.2    3
5.3    1
5.4    5
5.5    2
5.7    2
5.8    1
dtype: int64
```

In [6]:
```
SepalLengthprob = irisdf.groupby('SepallengthCm').size().div(len(irisdf))

SepalLengthprob
```

```
SepalLengthCm
4.3    0.02
4.4    0.06
4.5    0.02
4.6    0.08
4.7    0.04
4.8    0.10
4.9    0.08
5.0    0.16
5.1    0.16
5.2    0.06
5.3    0.02
5.4    0.10
5.5    0.04
5.7    0.04
5.8    0.02
dtype: float64
```

In [7]:
```
currsum = 0

for index, value in SepallengthprobIn.items():
    product = index * value
    currsum += product

print(f"Expected value : {currsum}")
```

```
Expected value : 5.005999999999999
```

In [8]:
```
sepallengthprobfrequency = pd.DataFrame({'SepallengthIncm':sepallengthfrequency.index, 'Frequency':sepallengthfrequency.values, 'Probability':SepalLengthprob.values})
sepallengthprobfrequency
```

Out[8]:

	SepalLengthIncm	Frequency	Probability
0	4.3	1	0.02
1	4.4	3	0.06
2	4.5	1	0.02
3	4.6	4	0.08
4	4.7	2	0.04
5	4.8	5	0.10
6	4.9	4	0.08
7	5.0	8	0.16
8	5.1	8	0.16
9	5.2	3	0.06
10	5.3	1	0.02
11	5.4	5	0.10
12	5.5	2	0.04
13	5.7	2	0.04
14	5.8	1	0.02

Post Lab

In [9]:
```
IrisVirginica = df.loc[df["Species"] == "Iris-virginica"]
IrisVirginica
```

Out[9]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
100	101	6.3	3.3	6.0	2.5	Iris-virginica
101	102	5.8	2.7	5.1	1.9	Iris-virginica
102	103	7.1	3.0	5.9	2.1	Iris-virginica
103	104	6.3	2.9	5.6	1.8	Iris-virginica
104	105	6.5	3.0	5.8	2.2	Iris-virginica
105	106	7.6	3.0	6.6	2.1	Iris-virginica
106	107	4.9	2.5	4.5	1.7	Iris-virginica
107	108	7.3	2.9	6.3	1.8	Iris-virginica
108	109	6.7	2.5	5.8	1.8	Iris-virginica
109	110	7.2	3.6	6.1	2.5	Iris-virginica
110	111	6.5	3.2	5.1	2.0	Iris-virginica
111	112	6.4	2.7	5.3	1.9	Iris-virginica
112	113	6.8	3.0	5.5	2.1	Iris-virginica
113	114	5.7	2.5	5.0	2.0	Iris-virginica
114	115	5.8	2.8	5.1	2.4	Iris-virginica
115	116	6.4	3.2	5.3	2.3	Iris-virginica
116	117	6.5	3.0	5.5	1.8	Iris-virginica
117	118	7.7	3.8	6.7	2.2	Iris-virginica
118	119	7.7	2.6	6.9	2.3	Iris-virginica
119	120	6.0	2.2	5.0	1.5	Iris-virginica
120	121	6.9	3.2	5.7	2.3	Iris-virginica
121	122	5.6	2.8	4.9	2.0	Iris-virginica
122	123	7.7	2.8	6.7	2.0	Iris-virginica
123	124	6.3	2.7	4.9	1.8	Iris-virginica
124	125	6.7	3.3	5.7	2.1	Iris-virginica
125	126	7.2	3.2	6.0	1.8	Iris-virginica
126	127	6.2	2.8	4.8	1.8	Iris-virginica
127	128	6.1	3.0	4.9	1.8	Iris-virginica
128	129	6.4	2.8	5.6	2.1	Iris-virginica
129	130	7.2	3.0	5.8	1.6	Iris-virginica
130	131	7.4	2.8	6.1	1.9	Iris-virginica
131	132	7.9	3.8	6.4	2.0	Iris-virginica
132	133	6.4	2.8	5.6	2.2	Iris-virginica
133	134	6.3	2.8	5.1	1.5	Iris-virginica
134	135	6.1	2.6	5.6	1.4	Iris-virginica
135	136	7.7	3.0	6.1	2.3	Iris-virginica
136	137	6.3	3.4	5.6	2.4	Iris-virginica
137	138	6.4	3.1	5.5	1.8	Iris-virginica
138	139	6.0	3.0	4.8	1.8	Iris-virginica
139	140	6.9	3.1	5.4	2.1	Iris-virginica
140	141	6.7	3.1	5.6	2.4	Iris-virginica
141	142	6.9	3.1	5.1	2.3	Iris-virginica
142	143	5.8	2.7	5.1	1.9	Iris-virginica
143	144	6.8	3.2	5.9	2.3	Iris-virginica
144	145	6.7	3.3	5.7	2.5	Iris-virginica
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

In [10]:
```
print("Unique values in Sepal Length CM are :", IrisVirginica['SepallengthCm'].unique())
```

```
Unique values in Sepal Length CM are : [6.3 5.8 7.1 6.5 7.6 4.9 7.3 6.7 7.2 6.4 6.8 5.7 7.7 6.  6.9 5.6 6.2 6.1
 7.4 7.9 5.9]
```

In [11]:
```
VirginicaSepallengthfrequency = IrisVirginica.groupby('SepallengthCm').size()

VirginicaSepallengthfrequency
```

```
SepalLengthCm
4.9    1
5.6    1
5.7    1
5.8    3
5.9    1
6.0    2
6.1    2
6.2    2
6.3    6
6.4    5
6.5    4
6.7    5
6.8    3
6.9    3
7.1    1
7.2    3
7.3    1
7.4    1
7.6    1
7.7    4
7.9    1
dtype: int64
```

In [12]:
```
virginicaSepallengthprob = IrisVirginica.groupby('SepallengthCm').size().div(len(irisdf))

virginicaSepallengthprob
```

```
SepalLengthCm
4.9    0.02
5.6    0.02
5.7    0.02
5.8    0.06
5.9    0.02
6.0    0.04
6.1    0.04
6.2    0.04
6.3    0.12
6.4    0.10
6.5    0.08
6.7    0.10
6.8    0.06
6.9    0.06
7.1    0.02
7.2    0.06
7.3    0.02
7.4    0.02
7.6    0.02
7.7    0.08
7.9    0.02
dtype: float64
```

In [13]:
```
virginicaSepallengthprobfrequency = pd.DataFrame({'SepallengthIncm':VirginicaSepallengthfrequency.index, 'Frequency':VirginicaSepallengthfrequency.values, 'Probability':virginicaSepallengthprob.values})
virginicaSepallengthprobfrequency
```

Out[13]:

	SepalLengthIncm	Frequency	Probability
0	4.9	1	0.02
1	5.6	1	0.02
2	5.7	1	0.02
3	5.8	3	0.06
4	5.9	1	0.02
5	6.0	2	0.04
6	6.1	2	0.04
7	6.2	2	0.04
8	6.3	6	0.12
9	6.4	5	0.10
10	6.5	4	0.08
11	6.7	5	0.10
12	6.8	3	0.06
13	6.9	3	0.06
14	7.1	1	0.02
15	7.2	3	0.06
16	7.3	1	0.02
17	7.4	1	0.02
18	7.6	1	0.02
19	7.7	4	0.08
20	7.9	1	0.02

In [14]:
```
currsum = 0

for index, value in virginicaSepallengthprob.items():
    product = index * value
    currsum += product

print(f"Expected value : {currsum}")
```

```
Expected value : 6.588
```