Assignment 3 Descriptive Statistics Iris Data

In [1]:
```
import pandas as pd
import seaborn as sns
import numpy as np
```

In [2]:
```
iris_data = sns.load_dataset('iris')
iris_data
```

Out[2]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

In [3]:
```
iris_data.shape
```

Out[3]:
```
(150, 5)
```

In [4]:
```
iris_data.species.unique()
```

Out[4]:
```
array(['setosa', 'versicolor', 'virginica'], dtype=object)
```

In [5]:
```
iris_data.species.nunique()
```

Out[5]:
```
3
```

In [6]:
```
iris_data.describe()
```

Out[6]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

In [7]:
```
iris_data.groupby(['species']).count()
```

Out[7]:

species	sepal_length	sepal_width	petal_length	petal_width
setosa	50	50	50	50
versicolor	50	50	50	50
virginica	50	50	50	50

In [8]:
```
setosa_data = iris_data[iris_data['species'] == 'setosa']
setosa_data
```

Out[8]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa
8	4.4	2.9	1.4	0.2	setosa
9	4.9	3.1	1.5	0.1	setosa
10	5.4	3.7	1.5	0.2	setosa
11	4.8	3.4	1.6	0.2	setosa
12	4.8	3.0	1.4	0.1	setosa
13	4.3	3.0	1.1	0.1	setosa
14	5.8	4.0	1.2	0.2	setosa
15	5.7	4.4	1.5	0.4	setosa
16	5.4	3.9	1.3	0.4	setosa
17	5.1	3.5	1.4	0.3	setosa
18	5.7	3.8	1.7	0.3	setosa
19	5.1	3.8	1.5	0.3	setosa
20	5.4	3.4	1.7	0.2	setosa
21	5.1	3.7	1.5	0.4	setosa
22	4.6	3.6	1.0	0.2	setosa
23	5.1	3.3	1.7	0.5	setosa
24	4.8	3.4	1.9	0.2	setosa
25	5.0	3.0	1.6	0.2	setosa
26	5.0	3.4	1.6	0.4	setosa
27	5.2	3.5	1.5	0.2	setosa
28	5.2	3.4	1.4	0.2	setosa

29	4.7	3.2	1.6	0.2	setosa
30	4.8	3.1	1.6	0.2	setosa
31	5.4	3.4	1.5	0.4	setosa
32	5.2	4.1	1.5	0.1	setosa
33	5.5	4.2	1.4	0.2	setosa
34	4.9	3.1	1.5	0.2	setosa
35	5.0	3.2	1.2	0.2	setosa
36	5.5	3.5	1.3	0.2	setosa
37	4.9	3.6	1.4	0.1	setosa
38	4.4	3.0	1.3	0.2	setosa
39	5.1	3.4	1.5	0.2	setosa
40	5.0	3.5	1.3	0.3	setosa
41	4.5	2.3	1.3	0.3	setosa
42	4.4	3.2	1.3	0.2	setosa
43	5.0	3.5	1.6	0.6	setosa
44	5.1	3.8	1.9	0.4	setosa
45	4.8	3.0	1.4	0.3	setosa
46	5.1	3.8	1.6	0.2	setosa
47	4.6	3.2	1.4	0.2	setosa
48	5.3	3.7	1.5	0.2	setosa
49	5.0	3.3	1.4	0.2	setosa

In [9]:
```
versicolor_data = iris_data[iris_data['species'] == 'versicolor']
versicolor_data
```

Out[9]:

	sepal_length	sepal_width	petal_length	petal_width	species
50	7.0	3.2	4.7	1.4	versicolor
51	6.4	3.2	4.5	1.5	versicolor
52	6.9	3.1	4.9	1.5	versicolor
53	5.5	2.3	4.0	1.3	versicolor
54	6.5	2.8	4.6	1.5	versicolor
55	5.7	2.8	4.5	1.3	versicolor
56	6.3	3.3	4.7	1.6	versicolor
57	4.9	2.4	3.3	1.0	versicolor
58	6.6	2.9	4.6	1.3	versicolor
59	5.2	2.7	3.9	1.4	versicolor
60	5.0	2.0	3.5	1.0	versicolor
61	5.9	3.0	4.2	1.5	versicolor
62	6.0	2.2	4.0	1.0	versicolor
63	6.1	2.9	4.7	1.4	versicolor
64	5.6	2.9	3.6	1.3	versicolor
65	6.7	3.1	4.4	1.4	versicolor

66	5.6	3.0	4.5	1.5	versicolor
67	5.8	2.7	4.1	1.0	versicolor
68	6.2	2.2	4.5	1.5	versicolor
69	5.6	2.5	3.9	1.1	versicolor
70	5.9	3.2	4.8	1.8	versicolor
71	6.1	2.8	4.0	1.3	versicolor
72	6.3	2.5	4.9	1.5	versicolor
73	6.1	2.8	4.7	1.2	versicolor
74	6.4	2.9	4.3	1.3	versicolor
75	6.6	3.0	4.4	1.4	versicolor
76	6.8	2.8	4.8	1.4	versicolor
77	6.7	3.0	5.0	1.7	versicolor
78	6.0	2.9	4.5	1.5	versicolor
79	5.7	2.6	3.5	1.0	versicolor
80	5.5	2.4	3.8	1.1	versicolor
81	5.5	2.4	3.7	1.0	versicolor
82	5.8	2.7	3.9	1.2	versicolor
83	6.0	2.7	5.1	1.6	versicolor
84	5.4	3.0	4.5	1.5	versicolor
85	6.0	3.4	4.5	1.6	versicolor
86	6.7	3.1	4.7	1.5	versicolor
87	6.3	2.3	4.4	1.3	versicolor
88	5.6	3.0	4.1	1.3	versicolor
89	5.5	2.5	4.0	1.3	versicolor
90	5.5	2.6	4.4	1.2	versicolor
91	6.1	3.0	4.6	1.4	versicolor
92	5.8	2.6	4.0	1.2	versicolor
93	5.0	2.3	3.3	1.0	versicolor
94	5.6	2.7	4.2	1.3	versicolor
95	5.7	3.0	4.2	1.2	versicolor
96	5.7	2.9	4.2	1.3	versicolor
97	6.2	2.9	4.3	1.3	versicolor
98	5.1	2.5	3.0	1.1	versicolor
99	5.7	2.8	4.1	1.3	versicolor

In [42]:
```
virginica_data = iris_data[iris_data['species'] == 'virginica']
virginica_data.head()
```

Out[42]:

	sepal_length	sepal_width	petal_length	petal_width	species
100	6.3	3.3	6.0	2.5	virginica
101	5.8	2.7	5.1	1.9	virginica
102	7.1	3.0	5.9	2.1	virginica

103	6.3	2.9	5.6	1.8	virginica
104	6.5	3.0	5.8	2.2	virginica

In [11]:
```
species_data_g = iris_data.groupby('species')
setosa_data_g
```

Out[11]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000022FBB29FA30>
```

In [29]:
```
setosa_data.describe()
```

Out[29]:

	sepal_length	sepal_width	petal_length	petal_width
count	50.00000	50.000000	50.000000	50.000000
mean	5.00600	3.428000	1.462000	0.246000
std	0.35249	0.379064	0.173664	0.105386
min	4.30000	2.300000	1.000000	0.100000
25%	4.80000	3.200000	1.400000	0.200000
50%	5.00000	3.400000	1.500000	0.200000
75%	5.20000	3.675000	1.575000	0.300000
max	5.80000	4.400000	1.900000	0.600000

In [13]:
```
versicolor_data.describe()
```

Out[13]:

	sepal_length	sepal_width	petal_length	petal_width
count	50.000000	50.000000	50.000000	50.000000
mean	5.936000	2.770000	4.260000	1.326000
std	0.516171	0.313798	0.469911	0.197753
min	4.900000	2.000000	3.000000	1.000000
25%	5.600000	2.525000	4.000000	1.200000
50%	5.900000	2.800000	4.350000	1.300000
75%	6.300000	3.000000	4.600000	1.500000
max	7.000000	3.400000	5.100000	1.800000

In [31]:
```
virginica_data.describe()
```

Out[31]:

	sepal_length	sepal_width	petal_length	petal_width
count	50.00000	50.000000	50.000000	50.00000
mean	6.58800	2.974000	5.552000	2.02600
std	0.63588	0.322497	0.551895	0.27465
min	4.90000	2.200000	4.500000	1.40000
25%	6.22500	2.800000	5.100000	1.80000
50%	6.50000	3.000000	5.550000	2.00000
75%	6.90000	3.175000	5.875000	2.30000
max	7.90000	3.800000	6.900000	2.50000

In [22]:
```
nc = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width','species']
# species_data = ['setosa','versicolor','viginica']
```

In [43]:
```
nc = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width']
def species_stats(species_data,species_name):
        print("Species Name: {}".format(species_name))
        print("Mean:",species_data[nc].mean())
        print("Median:",species_data[nc].median())
        print("std:",species_data[nc].std())
        print("25% percentile:",species_data[nc].quantile(0.25))
        print("75% percentile:",species_data[nc].quantile(0.75))
        print("Min:",species_data[nc].min())
        print("Max:",species_data[nc].max())
```

In [49]:
```
species_data_names = ['setosa_data','viginica_data','versicolor_data']
for data in species_data_names:
    print("************** Species name {} *****************".format(data))
    species_stats(setosa_data,data)
    print("-------------------------------------")
```

```
************** Species name setosa_data *****************
Species Name: setosa_data
Mean: sepal_length    5.006
sepal_width     3.428
petal_length    1.462
petal_width     0.246
dtype: float64
Median: sepal_length    5.0
sepal_width     3.4
petal_length    1.5
petal_width     0.2
dtype: float64
std: sepal_length    0.352490
sepal_width     0.379064
petal_length    0.173664
petal_width     0.105386
dtype: float64
25% percentile: sepal_length    4.8
sepal_width     3.2
petal_length    1.4
petal_width     0.2
Name: 0.25, dtype: float64
75% percentile: sepal_length    5.200
sepal_width     3.675
petal_length    1.575
petal_width     0.300
Name: 0.75, dtype: float64
Min: sepal_length    4.3
sepal_width     2.3
petal_length    1.0
petal_width     0.1
dtype: float64
Max: sepal_length    5.8
sepal_width     4.4
petal_length    1.9
petal_width     0.6
dtype: float64
-------------------------------------
************** Species name viginica_data *****************
Species Name: viginica_data
Mean: sepal_length    5.006
sepal_width     3.428
petal_length    1.462
petal_width     0.246
dtype: float64
```

```
Median: sepal_length    5.0
sepal_width     3.4
petal_length    1.5
petal_width     0.2
dtype: float64
std: sepal_length    0.352490
sepal_width     0.379064
petal_length    0.173664
petal_width     0.105386
dtype: float64
25% percentile: sepal_length    4.8
sepal_width     3.2
petal_length    1.4
petal_width     0.2
Name: 0.25, dtype: float64
75% percentile: sepal_length    5.200
sepal_width     3.675
petal_length    1.575
petal_width     0.300
Name: 0.75, dtype: float64
Min: sepal_length    4.3
sepal_width     2.3
petal_length    1.0
petal_width     0.1
dtype: float64
Max: sepal_length    5.8
sepal_width     4.4
petal_length    1.9
petal_width     0.6
dtype: float64
-------------------------------------
************** Species name versicolor_data *****************
Species Name: versicolor_data
Mean: sepal_length    5.006
sepal_width     3.428
petal_length    1.462
petal_width     0.246
dtype: float64
Median: sepal_length    5.0
sepal_width     3.4
petal_length    1.5
petal_width     0.2
dtype: float64
std: sepal_length    0.352490
sepal_width     0.379064
petal_length    0.173664
petal_width     0.105386
dtype: float64
25% percentile: sepal_length    4.8
sepal_width     3.2
petal_length    1.4
petal_width     0.2
Name: 0.25, dtype: float64
75% percentile: sepal_length    5.200
sepal_width     3.675
petal_length    1.575
petal_width     0.300
Name: 0.75, dtype: float64
Min: sepal_length    4.3
sepal_width     2.3
petal_length    1.0
petal_width     0.1
dtype: float64
Max: sepal_length    5.8
sepal_width     4.4
petal_length    1.9
```

```
petal_width     0.6
dtype: float64
-----------------------------------
```

In [41]:
```
print(setosa_data.nunique())
```

```
sepal_length    15
sepal_width     16
petal_length     9
petal_width      6
species          1
dtype: int64
```

In [52]:
```
def calculate_mean(data):
    if len(data)==0:
        return 0
    m = sum(data)/len(data)
    return m

def calculate_std(data,mean):
    if len(data)<=1:
        return 0
    difference_squared = sum((x-mean)**2 for x in data)
    ans = (difference_squared/(len(data)-1))**0.5
    return ans

def calculate_percentile(data,percentile):
    sorted_data = sorted(data)
    index = int(percentile*len(data))
    percentile_result = sorted_data[index]
    return percentile_result

def display_stats(species_data,species_name):
    column = nc
    print(f"\n***************Statistics for {species_name}*******************")

    # Mean
    mean_values = [calculate_mean(species_data[col]) for col in column]
    print("Mean: ")
    print(pd.Series(mean_values, index=column))

    # Standard Deviation
    std_values = [calculate_std(species_data[col],mean_values[i]) for i, col in enumerat
    print("\nStandard Deviation")
    print(pd.Series(std_values, index=column))

    # Percentile
    percentiles = [0.25, 0.75]
    for percentile_value in percentiles:
        percentile_values = [calculate_percentile(species_data[col], percentile_value) f
        print(f"\n{int(percentile_value * 100)}th Percentile : ")
        print(pd.Series(percentile_values, index=column))

display_stats(setosa_data, 'Iris-setosa')
display_stats(versicolor_data, 'Iris-versicolor')
display_stats(virginica_data, 'Iris-virginica')
```

```
***************Statistics for Iris-setosa*******************
Mean:
sepal_length    5.006
sepal_width     3.428
petal_length    1.462
petal_width     0.246
dtype: float64

Standard Deviation
```

```
sepal_length    0.352490
sepal_width     0.379064
petal_length    0.173664
petal_width     0.105386
dtype: float64

25th Percentile :
sepal_length    4.8
sepal_width     3.2
petal_length    1.4
petal_width     0.2
dtype: float64

75th Percentile :
sepal_length    5.2
sepal_width     3.7
petal_length    1.6
petal_width     0.3
dtype: float64

**********Statistics for Iris-versicolor**********
Mean:
sepal_length    5.936
sepal_width     2.770
petal_length    4.260
petal_width     1.326
dtype: float64

Standard Deviation
sepal_length    0.516171
sepal_width     0.313798
petal_length    0.469911
petal_width     0.197753
dtype: float64

25th Percentile :
sepal_length    5.6
sepal_width     2.5
petal_length    4.0
petal_width     1.2
dtype: float64

75th Percentile :
sepal_length    6.3
sepal_width     3.0
petal_length    4.6
petal_width     1.5
dtype: float64

**********Statistics for Iris-virginica**********
Mean:
sepal_length    6.588
sepal_width     2.974
petal_length    5.552
petal_width     2.026
dtype: float64

Standard Deviation
sepal_length    0.635880
sepal_width     0.322497
petal_length    0.551895
petal_width     0.274650
dtype: float64

25th Percentile :
sepal_length    6.2
```

```
sepal_width     2.8
petal_length    5.1
petal_width     1.8
dtype: float64

75th Percentile :
sepal_length    6.9
sepal_width     3.2
petal_length    5.9
petal_width     2.3
dtype: float64
```

Group By

In [ ]:
```
```

In [54]:
```
iris_data.groupby(["species"])["sepal_length"].mean()
```

Out[54]:
```
species
setosa        5.006
versicolor    5.936
virginica     6.588
Name: sepal_length, dtype: float64
```

In [55]:
```
iris_data.groupby(["species"])["sepal_length"].std()
```

Out[55]:
```
species
setosa        0.352490
versicolor    0.516171
virginica     0.635880
Name: sepal_length, dtype: float64
```

In [56]:
```
iris_data.groupby(["species"])["sepal_length"].describe()
```

Out[56]:

species	count	mean	std	min	25%	50%	75%	max
setosa	50.0	5.006	0.352490	4.3	4.800	5.0	5.2	5.8
versicolor	50.0	5.936	0.516171	4.9	5.600	5.9	6.3	7.0
virginica	50.0	6.588	0.635880	4.9	6.225	6.5	6.9	7.9

In [57]:
```
iris_data.groupby(["species"])["sepal_length"].quantile(q=0.25)
```

Out[57]:
```
species
setosa        4.800
versicolor    5.600
virginica     6.225
Name: sepal_length, dtype: float64
```

In [58]:
```
iris_data.groupby(["species"])["sepal_length"].quantile(q=0.75)
```

Out[58]:
```
species
setosa        5.2
versicolor    6.3
virginica     6.9
Name: sepal_length, dtype: float64
```

In [65]:
```
a=iris_data.groupby(["species"])["sepal_length"].mean()
print(a)
b=iris_data.groupby(["species"])["sepal_length"].median()
print(b)
```

```
list=[a,b]
print(list)
```

```
species
setosa        5.006
versicolor    5.936
virginica     6.588
Name: sepal_length, dtype: float64
species
setosa        5.0
versicolor    5.9
virginica     6.5
Name: sepal_length, dtype: float64
[species
setosa        5.006
versicolor    5.936
virginica     6.588
Name: sepal_length, dtype: float64, species
setosa        5.0
versicolor    5.9
virginica     6.5
Name: sepal_length, dtype: float64]
```

In [ ]: