```
In [ ]: import numpy as np
  import pandas as pd
  import seaborn as sns
  import matplotlib.pyplot as plt
```

# **Data Preparation**

```
In [ ]: df = sns.load_dataset('iris')
    df.head()
```

Out[ ]:		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

# **Central Tendency**

#### Mean

```
df.mean(numeric_only=True)
Out[]: sepal_length
                          5.843333
         sepal_width
                          3.057333
         petal_length
                          3.758000
         petal_width
                          1.199333
         dtype: float64
In [ ]: df.groupby("species").mean(numeric_only=True)
Out[]:
                    sepal_length sepal_width petal_length petal_width
           species
            setosa
                          5.006
                                       3.428
                                                    1.462
                                                                 0.246
         versicolor
                          5.936
                                       2.770
                                                    4.260
                                                                 1.326
          virginica
                          6.588
                                       2.974
                                                    5.552
                                                                 2.026
```

#### Median

```
df.median(numeric_only=True)
Out[]: sepal_length
                          5.80
         sepal_width
                          3.00
                          4.35
         petal length
         petal_width
                          1.30
         dtype: float64
In [ ]: df.groupby("species").median(numeric_only=True)
Out[]:
                    sepal_length sepal_width petal_length petal_width
           species
                             5.0
                                                                    0.2
                                          3.4
                                                      1.50
            setosa
         versicolor
                             5.9
                                          2.8
                                                      4.35
                                                                    1.3
          virginica
                                                      5.55
                                                                    2.0
                             6.5
                                          3.0
```

## **Dispersion**

#### **Variance**

```
In [ ]: df.var(numeric_only=True)
Out[ ]: sepal_length     0.685694
     sepal_width     0.189979
     petal_length     3.116278
     petal_width     0.581006
     dtype: float64
```

#### **Standard Deviation**

```
In [ ]: df.std(numeric_only=True)
Out[ ]: sepal_length     0.828066
     sepal_width     0.435866
     petal_length     1.765298
     petal_width     0.762238
     dtype: float64
```

### Relatedness

#### **Correlation**

In [ ]: df.corr(numeric\_only=True) Out[]: sepal\_length sepal\_width petal\_length petal\_width sepal\_length 1.000000 -0.117570 0.871754 0.817941 sepal\_width -0.117570 1.000000 -0.428440 -0.366126 petal\_length 0.871754 -0.428440 1.000000 0.962865 petal\_width 0.817941 -0.366126 0.962865 1.000000

#### **Covariance**

In [ ]: df.cov(numeric\_only=True)

Out[ ]:		sepal_length	sepal_width	petal_length	petal_width
	sepal_length	0.685694	-0.042434	1.274315	0.516271
	sepal_width	-0.042434	0.189979	-0.329656	-0.121639
	petal_length	1.274315	-0.329656	3.116278	1.295609
	petal_width	0.516271	-0.121639	1.295609	0.581006

# **Quantiles**

### **Quartiles**

In [ ]: df.quantile(np.arange(0.25, 1.1, 0.25), axis = 0, numeric\_only=True)

Out[ ]:		sepal_length	sepal_width	petal_length	petal_width
	0.25	5.1	2.8	1.60	0.3
	0.50	5.8	3.0	4.35	1.3
	0.75	6.4	3.3	5.10	1.8
	1.00	7.9	4.4	6.90	2.5

### **Deciles**

In [ ]: df.quantile(np.arange(0.1, 1.1, 0.1), axis = 0, numeric\_only=True)

Out[]:		sepal_length	sepal_width	petal_length	petal_width
	0.1	4.80	2.50	1.40	0.20
	0.2	5.00	2.70	1.50	0.20
	0.3	5.27	2.80	1.70	0.40
	0.4	5.60	3.00	3.90	1.16
	0.5	5.80	3.00	4.35	1.30
	0.6	6.10	3.10	4.64	1.50
	0.7	6.30	3.20	5.00	1.80
	8.0	6.52	3.40	5.32	1.90
	0.9	6.90	3.61	5.80	2.20
	1.0	7.90	4.40	6.90	2.50

### **Percentiles**

In [ ]: df.quantile(np.arange(0.01, 1.01, 0.01), axis = 0, numeric\_only=True)

Out[ ]:		sepal_length	sepal_width	petal_length	petal_width
	0.01	4.400	2.200	1.149	0.100
	0.02	4.400	2.200	1.200	0.100
	0.03	4.547	2.300	1.300	0.147
	0.04	4.600	2.300	1.300	0.200
	0.05	4.600	2.345	1.300	0.200
	•••				
	0.96	7.408	3.804	6.108	2.304
	0.97	7.653	3.900	6.353	2.400
	0.98	7.700	4.002	6.602	2.402
	0.99	7.700	4.151	6.700	2.500
	1.00	7.900	4.400	6.900	2.500

100 rows × 4 columns

### **Box Plot/Whisker Plot**

```
In [ ]: bdf = df.drop(['species'], axis=1)
bdf
```

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

150 rows × 4 columns

In [ ]: sns.boxplot(data=bdf)
 plt.show()

