

Practical No. 03

Descriptive Statistics - Measures of Central Tendency and variability Perform the following operations on any open source dataset (e.g., data.csv)

1. Provide summary statistics (mean, median, minimum, maximum, standard deviation) for a dataset (age, income etc.) with numeric variables grouped by one of the qualitative (categorical) variable. For example, if your categorical variable is age groups and quantitative variable is income, then provide summary statistics of income grouped by the age groups. Create a list that contains a numeric value for each response to the categorical variable.
2. Write a Python program to display some basic statistical details like percentile, mean, standard deviation etc. of the species of 'Iris-setosa', 'Iris-versicolor' and 'Iris- versicolor' of iris.csv dataset. Provide the codes with outputs and explain everything that you do in this step.

```
In [4]: import pandas as pd
```

```
In [6]: df = pd.read_csv("/home/kartik/Documents/Python Notebooks/Iris.csv")
```

```
In [7]: df
```

Out[7]:

	sepallength	sepalwidth	petallength	petalwidth	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 5 columns

```
In [8]: df.describe()
```

Out[8]:

	sepallength	sepalwidth	petallength	petalwidth
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
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 [10]: df["sepallength"].describe()

```
Out[10]: count    150.000000
          mean     5.843333
          std      0.828066
          min      4.300000
          25%     5.100000
          50%     5.800000
          75%     6.400000
          max      7.900000
          Name: sepallength, dtype: float64
```

In [11]: df.groupby("class").describe()

```
Out[11]:
```

	class	sepallength					sepalwidth					petallength					petalwidth					
		count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%
Iris-setosa	50.0	5.006	0.352490	4.3	4.800	5.0	5.2	5.8	50.0	3.418	0.433594	1.433594	4.3	4.900	5.7	6.9	50.0	1.462	0.198667	0.1	1.0	1.575
Iris-versicolor	50.0	5.936	0.516171	4.9	5.600	5.9	6.3	7.0	50.0	2.770	0.828066	1.328066	4.3	5.000	6.9	7.9	50.0	2.974	0.433594	0.2	4.600	
Iris-virginica	50.0	6.588	0.635880	4.9	6.225	6.5	6.9	7.9	50.0	2.974	0.433594	1.328066	4.3	5.700	7.9	9.0	50.0	3.464	0.198667	0.3	5.875	

3 rows × 32 columns



In [12]: df.groupby("class").describe().sum()

```
Out[12]:  sepallength  count    150.000000
              mean     17.530000
              std      1.504540
              min     14.100000
              25%    16.625000
              50%    17.400000
              75%    18.400000
              max     20.700000
            sepalwidth   count    150.000000
              mean     9.162000
              std      1.017319
              min     6.500000
              25%    8.450000
              50%    9.200000
              75%    9.850000
              max     11.600000
          petallength   count    150.000000
              mean    11.276000
              std      1.195317
              min     8.500000
              25%    10.500000
              50%    11.400000
              75%    12.050000
              max     13.900000
        petalwidth    count    150.000000
              mean    3.596000
              std      0.579612
              min     2.500000
              25%    3.200000
              50%    3.500000
              75%    4.100000
              max     4.900000
          dtype: float64
```

In []: