

ASSIGNMENT 3

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

```
In [3]: iris = sns.load_dataset('iris')
```

```
In [4]: iris.head()
```

```
Out[4]:
```

| | 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 |

```
In [5]: iris.describe()
```

Out[5]:

| | 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 [6]: `iris.describe(include='object')`

Out[6]:

| | species |
|---------------|---------|
| count | 150 |
| unique | 3 |
| top | setosa |
| freq | 50 |

In [7]: `iris_groupby = iris.groupby(by='species')`In [8]: `iris_groupby.std()`

Out[8]:

| | sepal_length | sepal_width | petal_length | petal_width |
|-------------------|--------------|-------------|--------------|-------------|
| species | | | | |
| setosa | 0.352490 | 0.379064 | 0.173664 | 0.105386 |
| versicolor | 0.516171 | 0.313798 | 0.469911 | 0.197753 |
| virginica | 0.635880 | 0.322497 | 0.551895 | 0.274650 |

In [9]: `iris_groupby.mean()`

Out[9]:

| | 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 |

```
In [10]: iris.groupby.median()
```

```
Out[10]:
```

| | sepal_length | sepal_width | petal_length | petal_width |
|------------|--------------|-------------|--------------|-------------|
| species | | | | |
| setosa | 5.0 | 3.4 | 1.50 | 0.2 |
| versicolor | 5.9 | 2.8 | 4.35 | 1.3 |
| virginica | 6.5 | 3.0 | 5.55 | 2.0 |

```
In [11]: iris.groupby.min()
```

```
Out[11]:
```

| | sepal_length | sepal_width | petal_length | petal_width |
|------------|--------------|-------------|--------------|-------------|
| species | | | | |
| setosa | 4.3 | 2.3 | 1.0 | 0.1 |
| versicolor | 4.9 | 2.0 | 3.0 | 1.0 |
| virginica | 4.9 | 2.2 | 4.5 | 1.4 |

```
In [12]: iris.groupby.max()
```

```
Out[12]:
```

| | sepal_length | sepal_width | petal_length | petal_width |
|------------|--------------|-------------|--------------|-------------|
| species | | | | |
| setosa | 5.8 | 4.4 | 1.9 | 0.6 |
| versicolor | 7.0 | 3.4 | 5.1 | 1.8 |
| virginica | 7.9 | 3.8 | 6.9 | 2.5 |

```
In [13]: iris.groupby.quantile()
```

```
Out[13]:
```

| | sepal_length | sepal_width | petal_length | petal_width |
|------------|--------------|-------------|--------------|-------------|
| species | | | | |
| setosa | 5.0 | 3.4 | 1.50 | 0.2 |
| versicolor | 5.9 | 2.8 | 4.35 | 1.3 |
| virginica | 6.5 | 3.0 | 5.55 | 2.0 |