Statistics

Mean

|  |  |
| --- | --- |
| Import | import statistics |
| Mean of a data (list) | statistics.mean(data) |
| Mean of array (numpy) | arr.mean() 🡪arr=data |
| Mean in Pandas | df.mean() 🡪df=data |
| Mean formula | mean = sum(data)/len(data) |

Median

|  |  |
| --- | --- |
| for List | statistics.median(data) |
| in numpy | np.median(data) |
| in padas | data[‘column\_name].median() |

Mode

|  |  |
| --- | --- |
| for list | statistics.mode(data) |
| in numpy | np.mode(a) 🡪 mode is not available in  numpy directly so import from scipy |
| in pandas | data[‘col\_name’].mode()[0] |

Standard Deviation

|  |  |
| --- | --- |
| in list | statistics.stdev(numbers) |
| in numpy | np.std(arr1) |
| in pandas | data['column'].std() |

Percentile

|  |  |
| --- | --- |
| in numpy | q1, q3 = np.percentile(arr1, [25,75] |
| qualntile menthod | a, b = np.quantile(arr1, [0.25, 0.75] |

Sorting

|  |  |
| --- | --- |
| sort normal list | dataset = sorted(dataset, reverse = False) |
| sort DataFrame in pandas | data.sort\_values(column\_name  ,axis=0,ascendeing=True) |

Random Sampling

|  |  |
| --- | --- |
| random sampling list | random.choice |
| in numpy | np.random.choice |
| import random library | import random |

T-Test

|  |  |
| --- | --- |
| Import | from scipy.stats import ttest\_1samp |
| test syntax | ttest, p\_value = ttest\_1samp(data\_sample,32)  here ttest and p\_values and data\_sample🡪 variables |
| Z -Score | z\_score = (i - mean)/std 🡪 i=value |