Day 2 - SQL - Measure of Dispersion_by_Virat Tiwari

October 8, 2023

1 Measure Of Dispersion

[1]:
```
age_list=[21,54,30,65,12,46,25,26,38,77,16,54,60,34,19,81,23,11,59,43,24,18]
```

Note - For getting mean (average) we should import numpy as np

[3]:
```
# WE CAN USE NUMPY FOR PERFORMING WIDE VARIETY OF MATHEMATICAL OPERATIONS OR␣
↪STATISTICAL OPERATIONS

import numpy as np
```

[4]:
```
np.mean(age_list)
```

[4]:
```
38.0
```

Variance - We can find the variance with the help of " np.var " function

[5]:
```
np.var(age_list)
```

[5]:
```
426.27272727272725
```

Standard Deviations - We can find the standard deviation with the help of " np.std " function

[6]:
```
np.std(age_list)
```

[6]:
```
20.646373223225606
```

Graphical Representation of distribution curve or distribution of data,

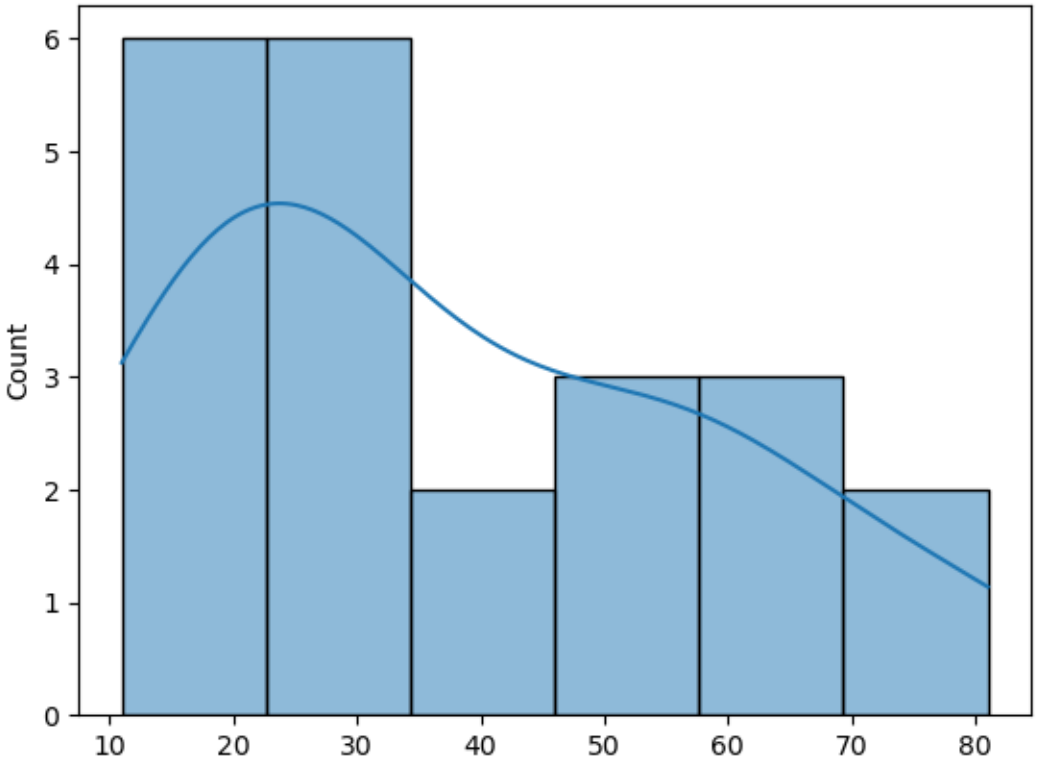
1 - First we import the seaborn as sns,

2- Second we use " sns.histplot(variable name , kde = True) " function for getting the graphical representation of data

[8]:
```
import seaborn as sns
sns.histplot(age_list,kde=True)

# THIS HISTOGRAM BASICALLY SHOWS THE SPREAD OR DISPERSION OF SPECIFIC DATA
# IN THIS HISTOGRAM VARIANCE IS LARGE SO SPREAD IS AUTOMATICALLY LARGE
```

[8]: <AxesSubplot: ylabel='Count'>



[9]:
```
#WE CAN USE PANDAS FOR DATA ANALYSIS AND IT MAKE CONCLUSION BASED ON␣
↪STATISTICAL THEORIES

import pandas as pd
```

[21]:
```
Data=[[45,36,45],[34,10,56],[18,31,19]]
```

[22]:
```
Data
```

[22]: [[45, 36, 45], [34, 10, 56], [18, 31, 19]]

[23]:
```
df=pd.DataFrame(Data,columns=["A","B","C"])
```

[24]:
```
df.head()
```

[24]:

	A	B	C
0	45	36	45
1	34	10	56
2	18	31	19

[25]:
```
df.var()
```

[25]:
```
A    184.333333
B    190.333333
C    361.000000
dtype: float64
```

[28]:
```
df.var(axis=1)
```

[28]:
```
0     27.000000
1    529.333333
2     52.333333
dtype: float64
```

[27]:
```
df.std()
```

[27]:
```
A    13.576941
B    13.796135
C    19.000000
dtype: float64
```

THANK YOU SO MUCH !!

YOURS VIRAT TIWARI :)