Deena ¶

20104016

libraries

In [1]:
```
import numpy as np
import pandas as pd
```

importing data set

In [2]:
```
df=pd.read_csv("4_drug200.csv")
df
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

200 rows × 6 columns

mean median mode()

In [3]:
```
df.mean()
```

Out[3]:
```
Age         44.315000
Na_to_K     16.084485
dtype: float64
```

In [4]:
```
df.median()
```

Out[4]:
```
Age        45.0000
Na_to_K    13.9365
dtype: float64
```

In [5]:
```
df.mode()
```

Out[5]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	47.0	M	HIGH	HIGH	12.006	drugY
1	NaN	NaN	NaN	NaN	18.295	NaN

describe ()

In [6]:
```
df.describe()
```

Out[6]:

	Age	Na_to_K
count	200.000000	200.000000
mean	44.315000	16.084485
std	16.544315	7.223956
min	15.000000	6.269000
25%	31.000000	10.445500
50%	45.000000	13.936500
75%	58.000000	19.380000
max	74.000000	38.247000

sum()

In [7]:
```
df.sum()
```

Out[7]:
```
Age                                                          8863
Sex            FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...
BP             HIGHLOWLOWNORMALLOWNORMALNORMALLOWNORMALLOWLOW...
Cholesterol    HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH...
Na_to_K                                                  3216.897
Drug           drugYdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
dtype: object
```

cumsum ()

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

Out[8]:

	Age	Sex	
0	23	F	
1	70	FM	
2	117	FMM	
3	145	FMMF	
4	206	FMMFF	
...	...	...	
195	8732	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...	HIGHLOWLOWNOR
196	8748	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...	HIGHLOWLOWNOR
197	8800	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...	HIGHLOWLOWNOR
198	8823	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...	HIGHLOWLOWNOR
199	8863	FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...	HIGHLOWLOWNOR

200 rows × 6 columns

min() and min()

In [9]:
```
df.min()
```

Out[9]:
```
Age              15
Sex               F
BP             HIGH
Cholesterol    HIGH
Na_to_K       6.269
Drug          drugA
dtype: object
```

In [10]:
```
df.max()
```

Out[10]:
```
Age              74
Sex               M
BP           NORMAL
Cholesterol  NORMAL
Na_to_K      38.247
Drug          drugY
dtype: object
```

count()

In [11]:
```
df.count()
```

Out[11]:
```
Age            200
Sex            200
BP             200
Cholesterol    200
Na_to_K        200
Drug           200
dtype: int64
```

Covariance

In [12]:
```
from numpy import cov
```

In [13]:
```
cov(df['Age'],df['Na_to_K'])
```

Out[13]:
```
array([[273.71434673,  -7.54375153],
       [ -7.54375153,  52.18553348]])
```

pearsonr and spearmanr

In [14]:
```
from scipy.stats import pearsonr
from scipy.stats import spearmanr
```

In [15]:
```
pearsonr(df['Age'],df['Na_to_K'])
```

Out[15]:
```
(-0.06311949726772592, 0.3745756399034559)
```

In [16]:
```
spearmanr(df['Age'],df['Na_to_K'])
```

Out[16]:
```
SpearmanrResult(correlation=-0.047273882688479915, pvalue=0.5062200581387418)
```