In [1]:
```
import numpy as np
import pandas as pd
from numpy import linalg as lg
from numpy import cov
```

5 Data Sets

In [2]:
```
a=pd.read_csv("drug.csv")
a
```

Out[2]:

	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

a) Find mean, median, mode and describe

In [3]:
```
print(a.mean)
print("Median:")
print(a.median)
print("Mode:")
print(a.mode)
print("Describe")
print(a.describe())
```

```
<bound method NDFrame._add_numeric_operations.<locals>.mean of      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 x 6 columns]>
Median:
<bound method NDFrame._add_numeric_operations.<locals>.median of      Age Sex      B
P 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 x 6 columns]>
Mode:
<bound method DataFrame.mode of      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 x 6 columns]>
Describe
              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
```

b) Find sum(), cumsum(), count, min and max values

In [4]:
```
print(a.sum())
```

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

In [5]:
```
print(a.cumsum())
```

```
      Age                                                  Sex  \
0      23                                                    F
1      70                                                   FM
2     117                                                  FMM
3     145                                                 FMMF
4     206                                                FMMFF
```

```
..    ...                                                  ...
195  8732  FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...
196  8748  FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...
197  8800  FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...
198  8823  FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...
199  8863  FMMFFFFMMMFFMFFFMMMFMMMFFFMFFMFMMFMMMMFMFFMMFF...

                                                       BP  \
0                                                    HIGH
1                                                  HIGHLOW
2                                                HIGHLOWLOW
3                                           HIGHLOWLOWNORMAL
4                                         HIGHLOWLOWNORMALLOW
..                                                       ...
195  HIGHLOWLOWNORMALLOWNORMALNORMALLOWNORMALLOWLOW...
196  HIGHLOWLOWNORMALLOWNORMALNORMALLOWNORMALLOWLOW...
197  HIGHLOWLOWNORMALLOWNORMALNORMALLOWNORMALLOWLOW...
198  HIGHLOWLOWNORMALLOWNORMALNORMALLOWNORMALLOWLOW...
199  HIGHLOWLOWNORMALLOWNORMALNORMALLOWNORMALLOWLOW...

                                              Cholesterol   Na_to_K  \
0                                                    HIGH    25.355
1                                                 HIGHHIGH    38.448
2                                              HIGHHIGHHIGH    48.562
3                                           HIGHHIGHHIGHHIGH    56.360
4                                         HIGHHIGHHIGHHIGHHIGH    74.403
..                                                       ...       ...
195  HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH...  3169.628
196  HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH...  3181.634
197  HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH...  3191.528
198  HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH...  3205.548
199  HIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHHIGHNORMALHIGH...  3216.897

                                                      Drug
0                                                    drugY
1                                                 drugYdrugC
2                                              drugYdrugCdrugC
3                                           drugYdrugCdrugCdrugX
4                                         drugYdrugCdrugCdrugXdrugY
..                                                       ...
195  drugYdrugCdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
196  drugYdrugCdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
197  drugYdrugCdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
198  drugYdrugCdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...
199  drugYdrugCdrugCdrugCdrugXdrugYdrugXdrugYdrugCdrugYd...

[200 rows x 6 columns]
```

In [6]:

```
print(a.count())
print(a.min())
print(a.max())
```

```
Age            200
Sex            200
BP             200
Cholesterol    200
Na_to_K        200
Drug           200
dtype: int64
Age              15
Sex               F
BP             HIGH
Cholesterol    HIGH
Na_to_K       6.269
Drug          drugA
dtype: object
Age              74
Sex               M
```

```
BP              NORMAL
Cholesterol     NORMAL
Na_to_K         38.247
Drug             drugY
dtype: object
```

c) Find covariance and correlation (spearman and pearsons)

In [7]:
```
print(a.cov())
```

```
                Age     Na_to_K
Age      273.714347  -7.543752
Na_to_K   -7.543752  52.185533
```

In [8]:
```
b=a.dropna(axis=1,how='any')
b
```

Out[8]:

	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

In [11]:
```
from scipy.stats import pearsonr
print(pearsonr(a['Age'],a['Na_to_K']))
```

```
(-0.06311949726772592, 0.3745756399034559)
```

In [13]:
```
from scipy.stats import spearmanr
print(spearmanr(a['Age'],a['Na_to_K']))
```

```
SpearmanrResult(correlation=-0.047273882688479915, pvalue=0.5062200581387418)
```

In [ ]:
```

```