statistical description

In [1]:
```
#Name : Achal Gajanan Ghorad
#Roll no. 39
#Section :3A
#Date:27/07/2024
```

In [2]:
```
# Aim: to perform data specialization
```

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

In [2]:
```
import os
```

In [3]:
```
os.getcwd()
```

Out[3]:
```
'C:\\Users\\ACHAL'
```

In [4]:
```
os.chdir("C:\\Users\\ACHAL\\OneDrive\\Desktop")
```

In [5]:
```
data=pd.read_csv("framingham.csv")
```

In [8]:
```
data.head()
```

Out[8]:

	Sex	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	preval
0	male	39	4.0	No	0.0	0.0	0	
1	female	46	2.0	No	0.0	0.0	0	
2	male	48	1.0	Yes	20.0	0.0	0	
3	female	61	3.0	Yes	30.0	0.0	0	
4	female	46	3.0	Yes	23.0	0.0	0	

In [9]:
```
1 data.head(100)
```

Out[9]:

	Sex	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	preva
0	male	39	4.0	No	0.0	0.0	0	
1	female	46	2.0	No	0.0	0.0	0	
2	male	48	1.0	Yes	20.0	0.0	0	
3	female	61	3.0	Yes	30.0	0.0	0	
4	female	46	3.0	Yes	23.0	0.0	0	
...	...	...	...	...	...	...	...	
95	female	65	3.0	No	0.0	0.0	0	
96	female	63	4.0	Yes	20.0	0.0	0	
97	female	40	2.0	No	0.0	0.0	0	
98	female	56	1.0	No	0.0	0.0	0	
99	female	56	1.0	Yes	15.0	0.0	0	

100 rows × 16 columns

In [10]:
```
1 data.tail()
```

Out[10]:

	Sex	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	pre
4235	female	48	2.0	Yes	20.0	NaN	0	
4236	female	44	1.0	Yes	15.0	0.0	0	
4237	female	52	2.0	No	0.0	0.0	0	
4238	male	40	3.0	No	0.0	0.0	0	
4239	female	39	3.0	Yes	30.0	0.0	0	

In [11]:
```
1 data.tail(10)
```

Out[11]:

	Sex	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	pre
4230	female	56	1.0	Yes	3.0	0.0	0	
4231	male	58	3.0	No	0.0	0.0	0	
4232	male	68	1.0	No	0.0	0.0	0	
4233	male	50	1.0	Yes	1.0	0.0	0	
4234	male	51	3.0	Yes	43.0	0.0	0	
4235	female	48	2.0	Yes	20.0	NaN	0	
4236	female	44	1.0	Yes	15.0	0.0	0	
4237	female	52	2.0	No	0.0	0.0	0	
4238	male	40	3.0	No	0.0	0.0	0	
4239	female	39	3.0	Yes	30.0	0.0	0	

In [12]:
```
1 data.describe()
```

Out[12]:

	age	education	cigsPerDay	BPMeds	prevalentStroke	prevalentHy
count	4240.000000	4135.000000	4211.000000	4187.000000	4240.000000	4240.00000
mean	49.580189	1.979444	9.005937	0.029615	0.005896	0.31061
std	8.572942	1.019791	11.922462	0.169544	0.076569	0.46279
min	32.000000	1.000000	0.000000	0.000000	0.000000	0.00000
25%	42.000000	1.000000	0.000000	0.000000	0.000000	0.00000
50%	49.000000	2.000000	0.000000	0.000000	0.000000	0.00000
75%	56.000000	3.000000	20.000000	0.000000	0.000000	1.00000
max	70.000000	4.000000	70.000000	1.000000	1.000000	1.00000

In [13]:
```
1 data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4240 entries, 0 to 4239
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   Sex              4240 non-null   object 
 1   age              4240 non-null   int64  
 2   education        4135 non-null   float64
 3   currentSmoker    4240 non-null   object 
 4   cigsPerDay       4211 non-null   float64
 5   BPMeds           4187 non-null   float64
 6   prevalentStroke  4240 non-null   int64  
 7   prevalentHyp     4240 non-null   int64  
 8   diabetes         4240 non-null   object 
 9   totChol          4190 non-null   float64
 10  sysBP            4240 non-null   float64
 11  diaBP            4240 non-null   float64
 12  BMI              4221 non-null   float64
 13  heartRate        4239 non-null   float64
 14  glucose          3852 non-null   float64
 15  TenYearCHD       4240 non-null   int64  
dtypes: float64(9), int64(4), object(3)
memory usage: 530.1+ KB
```

In [14]:
```
1 data.shape
```

Out[14]: (4240, 16)

In [6]:
```
1 data.size
```

Out[6]: 67840

In [7]:
```
1 data.ndim
```

Out[7]: 2

In [ ]:

```
1
```