

# Statistical Description

```
In [3]: #Name: Yash Pravin Gadbail  
#Roll no. : 35  
#Sec: 3rd A  
#Sub : ET 1  
#Date:27/07/2024
```

```
In [5]: # Aim: To Perform Statistical Description on Data
```

```
In [9]: import pandas as pd
```

```
In [11]: import os
```


```
In [15]: os.chdir("C:\\Users\\OneDrive\\Desktop")
```

```
In [17]: data=pd.read_csv("framingham.csv")
```

```
In [19]: data.head()
```

```
Out[19]:
```


	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
0	1	39	4.0	0	0.0	0.0	0	0
1	0	46	2.0	0	0.0	0.0	0	0
2	1	48	1.0	1	20.0	0.0	0	0
3	0	61	3.0	1	30.0	0.0	0	1
4	0	46	3.0	1	23.0	0.0	0	0



```
In [21]: data.head(10)
```

```
Out[21]:
```

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
0	1	39	4.0	0	0.0	0.0	0	0
1	0	46	2.0	0	0.0	0.0	0	0
2	1	48	1.0	1	20.0	0.0	0	0
3	0	61	3.0	1	30.0	0.0	0	1
4	0	46	3.0	1	23.0	0.0	0	0
5	0	43	2.0	0	0.0	0.0	0	1
6	0	63	1.0	0	0.0	0.0	0	0
7	0	45	2.0	1	20.0	0.0	0	0
8	1	52	1.0	0	0.0	0.0	0	1
9	1	43	1.0	1	30.0	0.0	0	1



In [23]: data.tail()

Out[23]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalent
4235	0	48	2.0	1	20.0	NaN	0	
4236	0	44	1.0	1	15.0	0.0	0	
4237	0	52	2.0	0	0.0	0.0	0	
4238	1	40	3.0	0	0.0	0.0	0	
4239	0	39	3.0	1	30.0	0.0	0	

In [27]: data.tail(10)

Out[27]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalent
4230	0	56	1.0	1	3.0	0.0	0	
4231	1	58	3.0	0	0.0	0.0	0	
4232	1	68	1.0	0	0.0	0.0	0	
4233	1	50	1.0	1	1.0	0.0	0	
4234	1	51	3.0	1	43.0	0.0	0	
4235	0	48	2.0	1	20.0	NaN	0	
4236	0	44	1.0	1	15.0	0.0	0	
4237	0	52	2.0	0	0.0	0.0	0	
4238	1	40	3.0	0	0.0	0.0	0	
4239	0	39	3.0	1	30.0	0.0	0	

In [31]: data.describe()

Out[31]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	pre
count	4240.000000	4240.000000	4135.000000	4240.000000	4211.000000	4187.000000	
mean	0.429245	49.580189	1.979444	0.494104	9.005937	0.029615	
std	0.495027	8.572942	1.019791	0.500024	11.922462	0.169544	
min	0.000000	32.000000	1.000000	0.000000	0.000000	0.000000	
25%	0.000000	42.000000	1.000000	0.000000	0.000000	0.000000	
50%	0.000000	49.000000	2.000000	0.000000	0.000000	0.000000	
75%	1.000000	56.000000	3.000000	1.000000	20.000000	0.000000	
max	1.000000	70.000000	4.000000	1.000000	70.000000	1.000000	

In [33]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4240 entries, 0 to 4239
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   male                   4240 non-null   int64  
1   age                    4240 non-null   int64  
2   education              4135 non-null   float64
3   currentSmoker          4240 non-null   int64  
4   cigsPerDay             4211 non-null   float64
5   BPMed                  4187 non-null   float64
6   prevalentStroke        4240 non-null   int64  
7   prevalentHyp           4240 non-null   int64  
8   diabetes               4240 non-null   int64  
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(7)
memory usage: 530.1 KB
```

In [35]: data.shape

Out[35]: (4240, 16)

In [37]: data.size

Out[37]: 67840

In [39]: data.ndim

Out[39]: 2