

Importing necessary modules

```
In [1]: 1 import pandas as pd
2 import numpy as np
3 import statistics as st
4 import matplotlib.pyplot as plt
5 import numpy as np
6
```

Importing data

```
In [2]: 1 car=pd.read_csv(r"K:\Desktop\NIIT\tables\Mtcars_Data_Practice.csv")
2 car
```

Out[2]:

	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Mazda RX4	21.0	6	160.0	210	3.90	2.620	16.46	0	1	4	4
1	Mazda RX4 Wag	21.0	6	160.0	210	3.90	2.875	17.02	0	1	4	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
3	Hornet 4 Drive	21.4	6	258.0	210	3.08	3.215	19.44	1	0	3	1
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
5	Valiant	18.1	6	225.0	205	2.76	3.460	20.22	1	0	3	1
6	Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
7	Merc 240D	24.4	4	146.7	330	3.69	3.190	20.00	1	0	4	2
8	Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
9	Merc 280	19.2	6	167.6	223	3.92	3.440	18.30	1	0	4	4
10	Merc 280C	17.8	6	167.6	223	3.92	3.440	18.90	1	0	4	4
11	Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
12	Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
13	Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
14	Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
15	Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
16	Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
17	Fiat 128	32.4	4	78.7	340	4.08	2.200	19.47	1	1	4	1
18	Honda Civic	30.4	4	75.7	280	4.93	1.615	18.52	1	1	4	2
19	Toyota Corolla	33.9	4	71.1	180	4.22	1.835	19.90	1	1	4	1
20	Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
21	Dodge Challenger	15.5	8	318.0	250	2.76	3.520	16.87	0	0	3	2
22	AMC Javelin	15.2	8	304.0	180	3.15	3.435	17.30	0	0	3	2
23	Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
24	Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
25	Fiat X1-9	27.3	4	79.0	290	4.08	1.935	18.90	1	1	4	1
26	Porsche 914-2	26.0	4	120.3	320	4.43	2.140	16.70	0	1	5	2
27	Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
28	Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
29	Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
30	Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
31	Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

Task1: Cars mt sampling and distribution analysis

Subtasks 3.1.1

```
In [3]: 1 rand_samp=car.sample(20,ignore_index=True,random_state=13)
2 rand_samp
```

Out[3]:

	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	AMC Javelin	15.2	8	304.0	180	3.15	3.435	17.30	0	0	3	2
1	Merc 240D	24.4	4	146.7	330	3.69	3.190	20.00	1	0	4	2
2	Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
3	Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
4	Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
5	Mazda RX4 Wag	21.0	6	160.0	210	3.90	2.875	17.02	0	1	4	4
6	Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
7	Fiat 128	32.4	4	78.7	340	4.08	2.200	19.47	1	1	4	1
8	Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
9	Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
10	Toyota Corolla	33.9	4	71.1	180	4.22	1.835	19.90	1	1	4	1
11	Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
12	Mazda RX4	21.0	6	160.0	210	3.90	2.620	16.46	0	1	4	4
13	Dodge Challenger	15.5	8	318.0	250	2.76	3.520	16.87	0	0	3	2
14	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
15	Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
16	Merc 280	19.2	6	167.6	223	3.92	3.440	18.30	1	0	4	4
17	Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
18	Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
19	Valiant	18.1	6	225.0	205	2.76	3.460	20.22	1	0	3	1

Subtasks 3.1.2

In [4]:

```
1 rand_samp_r=car.sample(20,replace=True,random_state=13,ignore_index=1)
2 rand_samp_r
```

Out[4]:

	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Honda Civic	30.4	4	75.7	280	4.93	1.615	18.52	1	1	4	2
1	Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
2	Merc 280C	17.8	6	167.6	223	3.92	3.440	18.90	1	0	4	4
3	Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
4	Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
5	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
6	Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
7	Fiat X1-9	27.3	4	79.0	290	4.08	1.935	18.90	1	1	4	1
8	Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
9	Porsche 914-2	26.0	4	120.3	320	4.43	2.140	16.70	0	1	5	2
10	Hornet 4 Drive	21.4	6	258.0	210	3.08	3.215	19.44	1	0	3	1
11	Porsche 914-2	26.0	4	120.3	320	4.43	2.140	16.70	0	1	5	2
12	Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
13	Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
14	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
15	Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
16	AMC Javelin	15.2	8	304.0	180	3.15	3.435	17.30	0	0	3	2
17	Valiant	18.1	6	225.0	205	2.76	3.460	20.22	1	0	3	1
18	Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
19	Merc 280C	17.8	6	167.6	223	3.92	3.440	18.90	1	0	4	4

Subtasks 3.1.3

In [5]:

```
1 sys_samp=car.iloc[np.arange(0,len(car),step=2)]
2 sys_samp
```

Out[5]:

	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Mazda RX4	21.0	6	160.0	210	3.90	2.620	16.46	0	1	4	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
6	Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
8	Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
10	Merc 280C	17.8	6	167.6	223	3.92	3.440	18.90	1	0	4	4
12	Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
14	Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
16	Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
18	Honda Civic	30.4	4	75.7	280	4.93	1.615	18.52	1	1	4	2
20	Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
22	AMC Javelin	15.2	8	304.0	180	3.15	3.435	17.30	0	0	3	2
24	Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
26	Porsche 914-2	26.0	4	120.3	320	4.43	2.140	16.70	0	1	5	2
28	Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
30	Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8

Subtasks 3.1.4

In [29]:

```
1 len(car.groupby("vs").get_group(1))
```

Out[29]: 14

In [28]:

```
1 strat_samp=pd.DataFrame()
2 samp_v=car.groupby("vs").get_group(1).sample(10,replace=True,ignore_index=True)
3 samp_s=car.groupby("vs").get_group(0).sample(10,replace=True,ignore_index=True)
4 strat_samp=pd.concat([strat_samp,samp_s,samp_v],ignore_index=True)
5 strat_samp
```

Out[28]:

	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
1	Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
2	Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
3	Porsche 914-2	26.0	4	120.3	320	4.43	2.140	16.70	0	1	5	2
4	Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
5	Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
6	Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
7	Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
8	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
9	Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
10	Merc 280	19.2	6	167.6	223	3.92	3.440	18.30	1	0	4	4
11	Merc 240D	24.4	4	146.7	330	3.69	3.190	20.00	1	0	4	2
12	Toyota Corolla	33.9	4	71.1	180	4.22	1.835	19.90	1	1	4	1
13	Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
14	Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
15	Hornet 4 Drive	21.4	6	258.0	210	3.08	3.215	19.44	1	0	3	1
16	Hornet 4 Drive	21.4	6	258.0	210	3.08	3.215	19.44	1	0	3	1
17	Merc 240D	24.4	4	146.7	330	3.69	3.190	20.00	1	0	4	2
18	Honda Civic	30.4	4	75.7	280	4.93	1.615	18.52	1	1	4	2
19	Merc 240D	24.4	4	146.7	330	3.69	3.190	20.00	1	0	4	2

Subtasks 3.1.5

In [31]:

```
1 clust_samp=pd.DataFrame()  
2 samp_v=car.groupby("vs").get_group(1).sample(frac=0.7,replace=True,ignore_index=True,random_state=13)  
3 samp_s=car.groupby("vs").get_group(0).sample(frac=0.6,replace=True,ignore_index=True,random_state=13)  
4 clust_samp=pd.concat([clust_samp,samp_s,samp_v],ignore_index=True)  
5 clust_samp
```

Out[31]:

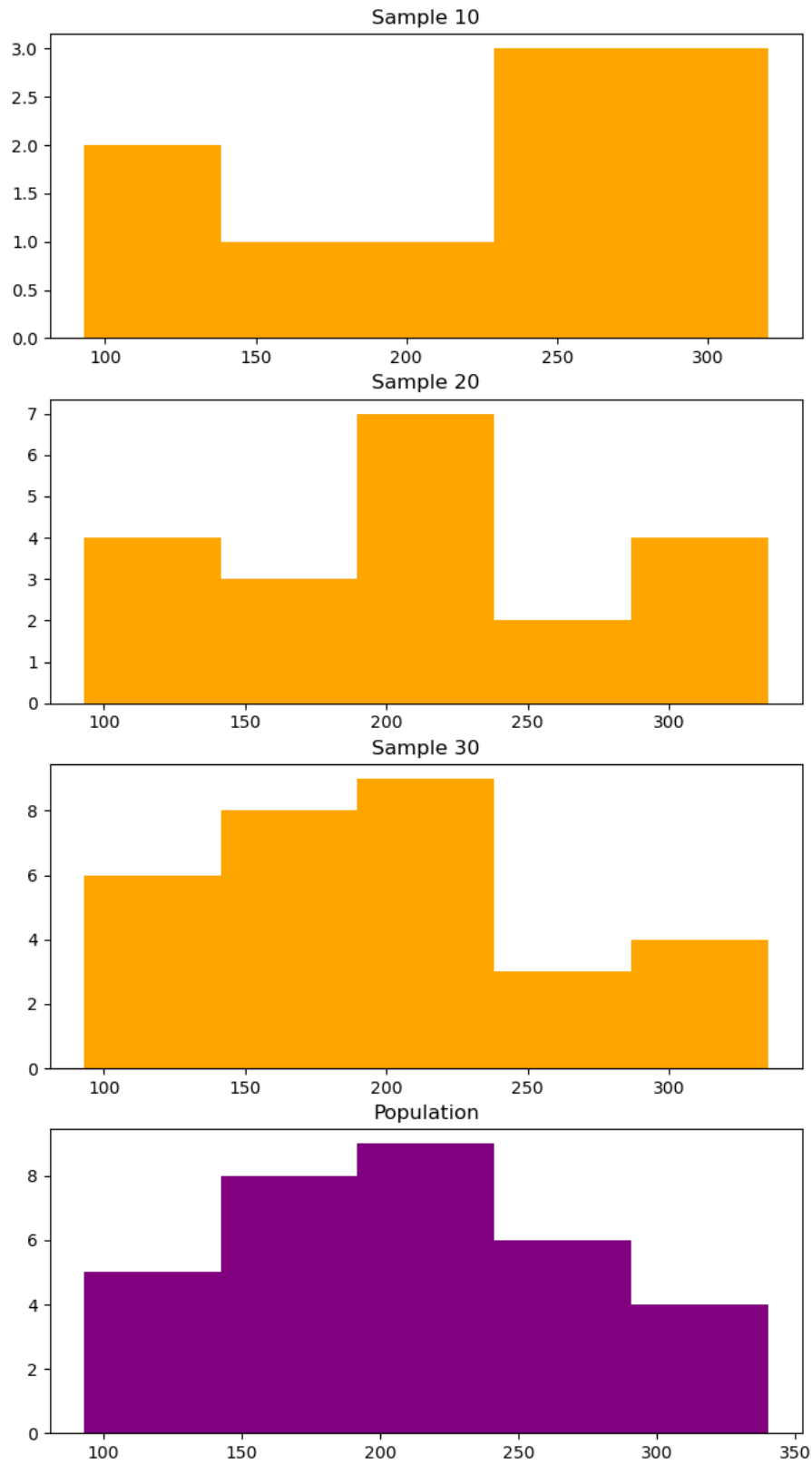
	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
1	Dodge Challenger	15.5	8	318.0	250	2.76	3.520	16.87	0	0	3	2
2	Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
3	Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
5	Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
6	Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
7	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
8	Porsche 914-2	26.0	4	120.3	320	4.43	2.140	16.70	0	1	5	2
9	Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
10	Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
11	Valiant	18.1	6	225.0	205	2.76	3.460	20.22	1	0	3	1
12	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
13	Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
14	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
15	Merc 280C	17.8	6	167.6	223	3.92	3.440	18.90	1	0	4	4
16	Valiant	18.1	6	225.0	205	2.76	3.460	20.22	1	0	3	1
17	Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
18	Toyota Corolla	33.9	4	71.1	180	4.22	1.835	19.90	1	1	4	1
19	Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
20	Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1

Task2: To find measures of dispersion and display them visually for 3 samples

```
In [56]: 1 fig,ax=plt.subplots(4,1,figsize=(8,15))
2
3 mean=[]
4 median=[]
5 std=[]
6 for i in range(1,5):
7     if(i==4):
8         mean.append(car.hp.mean())
9         median.append(car.hp.median())
10        std.append(car.hp.std())
11        ax[3].hist(car.hp,bins=5,color="purple")
12        ax[3].set_title("Population")
13        measures=pd.DataFrame([mean,median,std],columns=["Samp_10","Samp_20","Samp_30","Population"],index=["Mean","Median","Standard_dev"])
14    else:
15        samp=car.sample(i*10,replace=True,ignore_index=True,random_state=13)
16        ax[i-1].hist(samp.hp,bins=5,color="orange")
17        ax[i-1].set_title("Sample "+str(i*10))
18        mean.append(samp.hp.mean())
19        median.append(samp.hp.median())
20        std.append(samp.hp.std())
21 measures
22
```

Out[56]:

	Samp_10	Samp_20	Samp_30	Population
Mean	218.800000	211.800000	200.30000	211.312500
Median	230.000000	216.500000	205.00000	210.000000
Standard_dev	76.141389	75.772929	69.24676	68.513449



Interpretation

```
In [61]: 1 """From the above population graph its clear that it has gaussian distribution and the data is spread evenly but when looking at the sampled data we can see that for each random sample
2 that the mean and median are very close each other and standard deviation is similar with all the samples indicating proper
3 spread throughout the data"""
```

Out[61]: 'From the above population graph its clear that it has gaussian distribution and the data is spread evenly but when looking at the sampled data we can see that for each random sample the spread is uneven and unpredicatable but the observable point is the more the data points the more better the histogram follows gaussian distribution, also from measures of dispersion we can see \nthat the mean and median are very close each other and standard deviation is similar with all the samples indicating proper\nspread throughout the data'