

```
In [1]: 1 pip install numpy
        2
```

Requirement already satisfied: numpy in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (1.24.3)  
Note: you may need to restart the kernel to use updated packages.

```
In [2]: 1 pip install pip --upgrade
```

Requirement already satisfied: pip in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (23.1.2)Note: you may need to restart the kernel to use updated packages.

```
In [3]: 1 pip install numpy
        2
```

Requirement already satisfied: numpy in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (1.24.3)  
Note: you may need to restart the kernel to use updated packages.

```
In [4]: 1 pip install pandas
        2
```

Requirement already satisfied: pandas in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (2.0.1)Note: you may need to restart the kernel to use updated packages.

Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from pandas) (2.8.2)  
Requirement already satisfied: pytz>=2020.1 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from pandas) (2023.3)  
Requirement already satisfied: tzdata>=2022.1 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from pandas) (2023.3)  
Requirement already satisfied: numpy>=1.21.0 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from pandas) (1.24.3)  
Requirement already satisfied: six>=1.5 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)

```
In [5]: 1 import numpy as np
        2 import pandas as pd
```

```
In [6]: 1 df=pd.read_csv(r"C:\Users\teppa\Downloads\fiat500_VehicleSelection_Dataset")
2 df
```

Out[6]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...
342	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
343	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
344	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
345	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
346	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

347 rows × 112 columns

```
In [7]: 1 df.head(10)
```

Out[7]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	U
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

10 rows × 112 columns

```
In [8]: 1 df.shape
```

Out[8]: (347, 112)

```
In [9]: 1 df.describe()
```

Out[9]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
<b>count</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>mean</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>std</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>min</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>25%</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>50%</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>75%</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>max</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

8 rows × 112 columns



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

Out[10]:

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
<b>342</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>343</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>344</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>345</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
<b>346</b>	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 112 columns



In [11]:

```
1 df.isna
```

```
Out[11]: <bound method DataFrame.isna of
med: 3  Unnamed: 4  Unnamed: 5
0      NaN      NaN      NaN      NaN      NaN      NaN
\
1      NaN      NaN      NaN      NaN      NaN      NaN
2      NaN      NaN      NaN      NaN      NaN      NaN
3      NaN      NaN      NaN      NaN      NaN      NaN
4      NaN      NaN      NaN      NaN      NaN      NaN
..      ...      ...      ...      ...      ...      ...
342     NaN      NaN      NaN      NaN      NaN      NaN
343     NaN      NaN      NaN      NaN      NaN      NaN
344     NaN      NaN      NaN      NaN      NaN      NaN
345     NaN      NaN      NaN      NaN      NaN      NaN
346     NaN      NaN      NaN      NaN      NaN      NaN
```

```
      Unnamed: 6  Unnamed: 7  Unnamed: 8  Unnamed: 9  ...  Unnamed: 102
0      NaN      NaN      NaN      NaN      ...      NaN  \
1      NaN      NaN      NaN      NaN      ...      NaN
2      NaN      NaN      NaN      NaN      ...      NaN
3      NaN      NaN      NaN      NaN      ...      NaN
4      NaN      NaN      NaN      NaN      ...      NaN
..      ...      ...      ...      ...      ...      ...
342     NaN      NaN      NaN      NaN      ...      NaN
343     NaN      NaN      NaN      NaN      ...      NaN
344     NaN      NaN      NaN      NaN      ...      NaN
345     NaN      NaN      NaN      NaN      ...      NaN
346     NaN      NaN      NaN      NaN      ...      NaN
```

```
      Unnamed: 103  Unnamed: 104  Unnamed: 105  Unnamed: 106  Unnamed: 107
0      NaN      NaN      NaN      NaN      NaN  \
1      NaN      NaN      NaN      NaN      NaN
2      NaN      NaN      NaN      NaN      NaN
3      NaN      NaN      NaN      NaN      NaN
4      NaN      NaN      NaN      NaN      NaN
..      ...      ...      ...      ...      ...
342     NaN      NaN      NaN      NaN      NaN
343     NaN      NaN      NaN      NaN      NaN
344     NaN      NaN      NaN      NaN      NaN
345     NaN      NaN      NaN      NaN      NaN
346     NaN      NaN      NaN      NaN      NaN
```

```
      Unnamed: 108  Unnamed: 109  Unnamed: 110  Unnamed: 111
0      NaN      NaN      NaN      NaN
1      NaN      NaN      NaN      NaN
2      NaN      NaN      NaN      NaN
3      NaN      NaN      NaN      NaN
4      NaN      NaN      NaN      NaN
..      ...      ...      ...      ...
342     NaN      NaN      NaN      NaN
343     NaN      NaN      NaN      NaN
344     NaN      NaN      NaN      NaN
345     NaN      NaN      NaN      NaN
346     NaN      NaN      NaN      NaN
```

```
[347 rows x 112 columns]>
```

```
In [12]: 1 df.isna().any()  
        2
```

```
Out[12]: Unnamed: 0      True  
         Unnamed: 1      True  
         Unnamed: 2      True  
         Unnamed: 3      True  
         Unnamed: 4      True  
         ...  
         Unnamed: 107     True  
         Unnamed: 108     True  
         Unnamed: 109     True  
         Unnamed: 110     True  
         Unnamed: 111     True  
         Length: 112, dtype: bool
```

In [13]:

```
1 df.info
```

```

Out[13]: <bound method DataFrame.info of
med: 3  Unnamed: 4  Unnamed: 5
0      NaN      NaN      NaN      NaN      NaN      NaN
\
1      NaN      NaN      NaN      NaN      NaN      NaN
2      NaN      NaN      NaN      NaN      NaN      NaN
3      NaN      NaN      NaN      NaN      NaN      NaN
4      NaN      NaN      NaN      NaN      NaN      NaN
..      ...      ...      ...      ...      ...      ...
342     NaN      NaN      NaN      NaN      NaN      NaN
343     NaN      NaN      NaN      NaN      NaN      NaN
344     NaN      NaN      NaN      NaN      NaN      NaN
345     NaN      NaN      NaN      NaN      NaN      NaN
346     NaN      NaN      NaN      NaN      NaN      NaN

```

```

      Unnamed: 6  Unnamed: 7  Unnamed: 8  Unnamed: 9  ...  Unnamed: 102
0      NaN      NaN      NaN      NaN      ...      NaN  \
1      NaN      NaN      NaN      NaN      ...      NaN
2      NaN      NaN      NaN      NaN      ...      NaN
3      NaN      NaN      NaN      NaN      ...      NaN
4      NaN      NaN      NaN      NaN      ...      NaN
..      ...      ...      ...      ...      ...      ...
342     NaN      NaN      NaN      NaN      ...      NaN
343     NaN      NaN      NaN      NaN      ...      NaN
344     NaN      NaN      NaN      NaN      ...      NaN
345     NaN      NaN      NaN      NaN      ...      NaN
346     NaN      NaN      NaN      NaN      ...      NaN

```

```

      Unnamed: 103  Unnamed: 104  Unnamed: 105  Unnamed: 106  Unnamed: 107
0      NaN      NaN      NaN      NaN      NaN  \
1      NaN      NaN      NaN      NaN      NaN
2      NaN      NaN      NaN      NaN      NaN
3      NaN      NaN      NaN      NaN      NaN
4      NaN      NaN      NaN      NaN      NaN
..      ...      ...      ...      ...      ...
342     NaN      NaN      NaN      NaN      NaN
343     NaN      NaN      NaN      NaN      NaN
344     NaN      NaN      NaN      NaN      NaN
345     NaN      NaN      NaN      NaN      NaN
346     NaN      NaN      NaN      NaN      NaN

```

```

      Unnamed: 108  Unnamed: 109  Unnamed: 110  Unnamed: 111
0      NaN      NaN      NaN      NaN
1      NaN      NaN      NaN      NaN
2      NaN      NaN      NaN      NaN
3      NaN      NaN      NaN      NaN
4      NaN      NaN      NaN      NaN
..      ...      ...      ...      ...
342     NaN      NaN      NaN      NaN
343     NaN      NaN      NaN      NaN
344     NaN      NaN      NaN      NaN
345     NaN      NaN      NaN      NaN
346     NaN      NaN      NaN      NaN

```

[347 rows x 112 columns]>



In [14]: 1 pip install scipy

Requirement already satisfied: scipy in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (1.10.1)  
Requirement already satisfied: numpy<1.27.0,>=1.19.5 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from scipy) (1.24.3)  
Note: you may need to restart the kernel to use updated packages.

In [15]: 1 import numpy as np  
2 import pandas as pd  
3 from numpy.random import randn  
4 from numpy.random import seed  
5 from scipy.stats import spearmanr  
6 seed(1)  
7 data1=20\*randn(1000)+100  
8 data2=data1+(0\*randn(1000)+50)  
9 corr,\_=spearmanr(data1,data2)  
10 print('spearman correlation%.3f' %corr)

spearman correlation1.000

In [16]: 1 import numpy as np  
2 import pandas as pd  
3 from numpy.random import randn  
4 from numpy.random import seed  
5 from numpy import cov  
6 seed(1)  
7 data1=20\*randn(1000)+100  
8 data2=data1+(10\*randn(1000)+50)  
9 covariance=cov(data1,data2)  
10 print(covariance)

[[385.33297729 389.7545618 ]  
 [389.7545618 500.38006058]]

In [17]: 1 import numpy as np  
2 import pandas as pd  
3 from numpy.random import randn  
4 from numpy.random import seed  
5 from scipy.stats import pearsonr  
6 data1=20\*randn(1000)+100  
7 data2=data1+(10\*randn(1000)+50)  
8 corr,\_=pearsonr(data1,data2)  
9 print("pearson correlation:%.3f" %corr)

pearson correlation:0.891

In [18]: 1 pip install statsmodels

```
Requirement already satisfied: statsmodels in c:\users\teppa\appdata\local\pr  
ograms\python\python310\lib\site-packages (0.14.0)  
Requirement already satisfied: numpy>=1.18 in c:\users\teppa\appdata\local\pr  
ograms\python\python310\lib\site-packages (from statsmodels) (1.24.3)  
Requirement already satisfied: scipy!=1.9.2,>=1.4 in c:\users\teppa\appdata\l  
ocal\programs\python\python310\lib\site-packages (from statsmodels) (1.10.1)  
Requirement already satisfied: pandas>=1.0 in c:\users\teppa\appdata\local\pr  
ograms\python\python310\lib\site-packages (from statsmodels) (2.0.1)  
Requirement already satisfied: patsy>=0.5.2 in c:\users\teppa\appdata\local\pr  
ograms\python\python310\lib\site-packages (from statsmodels) (0.5.3)  
Requirement already satisfied: packaging>=21.3 in c:\users\teppa\appdata\loca  
l\programs\python\python310\lib\site-packages (from statsmodels) (23.1)  
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\teppa\appda  
ta\local\programs\python\python310\lib\site-packages (from pandas>=1.0->stats  
models) (2.8.2)  
Requirement already satisfied: pytz>=2020.1 in c:\users\teppa\appdata\local\p  
rograms\python\python310\lib\site-packages (from pandas>=1.0->statsmodels) (2  
023.3)  
Requirement already satisfied: tzdata>=2022.1 in c:\users\teppa\appdata\local  
\programs\python\python310\lib\site-packages (from pandas>=1.0->statsmodels)  
(2023.3)  
Requirement already satisfied: six in c:\users\teppa\appdata\local\programs\p  
ython\python310\lib\site-packages (from patsy>=0.5.2->statsmodels) (1.16.0)  
Note: you may need to restart the kernel to use updated packages.
```

In [19]:

```
1 import pandas as pd  
2 import statsmodels as sm  
3 from scipy import stats  
4 from statsmodels.stats import weightstats as stests  
5 data=[89,93,97,98,96,99,97,110,104,119,105,104,110,110,112,115,114]  
6 z_test,p_val=stests.ztest(data,x2=None,value=160)  
7 print(p_val)  
8 if p_val<0.05:  
9     print("We can reject the null hypothesis")  
10 else:  
11     print("we can accet the null hypothesis")  
12  
13
```

5.5796197354978736e-157

We can reject the null hypothesis

In [20]:

```
1 from scipy.stats import ttest_1samp
2 import numpy as np
3 ages=[45,89,23,46,12,69,45,24,34,67]
4 print(ages)
5 mean=np.mean(ages)
6 print(mean)
7 t_test,p_val=ttest_1samp(ages,30)
8 print("_value is :",p_val)
9 if p_val<0.5:
10     print("we can reject the null hypothesis")
11 else:
12     print("We can accept the null hypothesis")
13
```

[45, 89, 23, 46, 12, 69, 45, 24, 34, 67]

45.4

\_value is : 0.07179988272763561

we can reject the null hypothesis

In [21]:

```
1 from scipy.stats import ttest_ind
2 import numpy as np
3 data_group1=np.array([12,18,12,13,15,1,7,20,21,25,19,31,21,17,15,19,5,12,
4 data_group2=np.array([23,22,24,25,21,26,21,21,25,30,24,21,23,19,14,18,14,
5 mean1=np.mean(data_group1)
6 mean2=np.mean(data_group2)
7 print("Data group1.mean value:",mean1)
8 print("Data group2.mean value:",mean2)
9 std1=np.std(data_group1)
10 std2=np.std(data_group2)
11 print("Data group1 std value:",std1)
12 print("Data group2 std value :",std2)
13 t_test,p_val=ttest_ind(data_group1,data_group2)
14 print("The p_value is :",p_val)
15 if p_val<0.5:
16     print("we can reject the null hypothesis")
17 else:
18     print("We can accept the null hypothesis")
```

Data group1.mean value: 15.68421052631579

Data group2.mean value: 21.35

Data group1 std value: 6.805977493342834

Data group2 std value : 4.327528162819972

The p\_value is : 0.00436107719276935

we can reject the null hypothesis

```
In [22]: 1 import scipy.stats
2 data1=[0.0842,0.0368,0.0847,0.0935,0.0376,0.0963,0.0684,0.0758,0.0854,0.0
3 data2=[0.0785,0.0845,0.0758,0.0853,0.0946,0.0785,0.0853,0.0685]
4 data3=[0.0864,0.2522,0.0894,0.2724,0.0853,0.1367,0.853]
5 F_test,p_val=scipy.stats.f_oneway(data1,data2,data3)
6 print("p_value is:",p_val)
7 if p_val<0.5:
8     print("we can reject the null hypothesis")
9 else:
10     print("We can accept the null hypothesis")
```

p\_value is: 0.04043792126789142  
we can reject the null hypothesis

```
In [23]: 1 from scipy.stats import chi2_contingency
2 data=[[231,256,321],[245,312,213]]
3 test,p_val,dof,expected_val=chi2_contingency(data)
4 alpha=0.05
5 print("The p_value of our test is",((p_val)))
6 if p_val<=alpha:
7     print("we can reject")
8 else:
9     print("we can accept the null hypothesis")
```

The p\_value of our test is 1.4585823594475804e-06  
we can reject

```
In [24]: 1 pip install scikit-learn
```

Requirement already satisfied: scikit-learn in c:\users\teppa\appdata\local\p  
rograms\python\python310\lib\site-packages (1.2.2)  
Requirement already satisfied: numpy>=1.17.3 in c:\users\teppa\appdata\local  
\programs\python\python310\lib\site-packages (from scikit-learn) (1.24.3)  
Requirement already satisfied: scipy>=1.3.2 in c:\users\teppa\appdata\local\p  
rograms\python\python310\lib\site-packages (from scikit-learn) (1.10.1)  
Requirement already satisfied: joblib>=1.1.1 in c:\users\teppa\appdata\local  
\programs\python\python310\lib\site-packages (from scikit-learn) (1.2.0)  
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\teppa\appdata  
\local\programs\python\python310\lib\site-packages (from scikit-learn) (3.1.  
0)  
Note: you may need to restart the kernel to use updated packages.

```
In [25]: 1 from sklearn.metrics import mean_squared_error
2 Y_act=[1,4,3,2,6]
3 Y_pred=[0.6,1.29,2.69,3.4,5.2]
4 mean_squared_error(Y_act,Y_pred)
```

Out[25]: 2.04004

```
In [26]: 1 from scipy.stats import binom
```

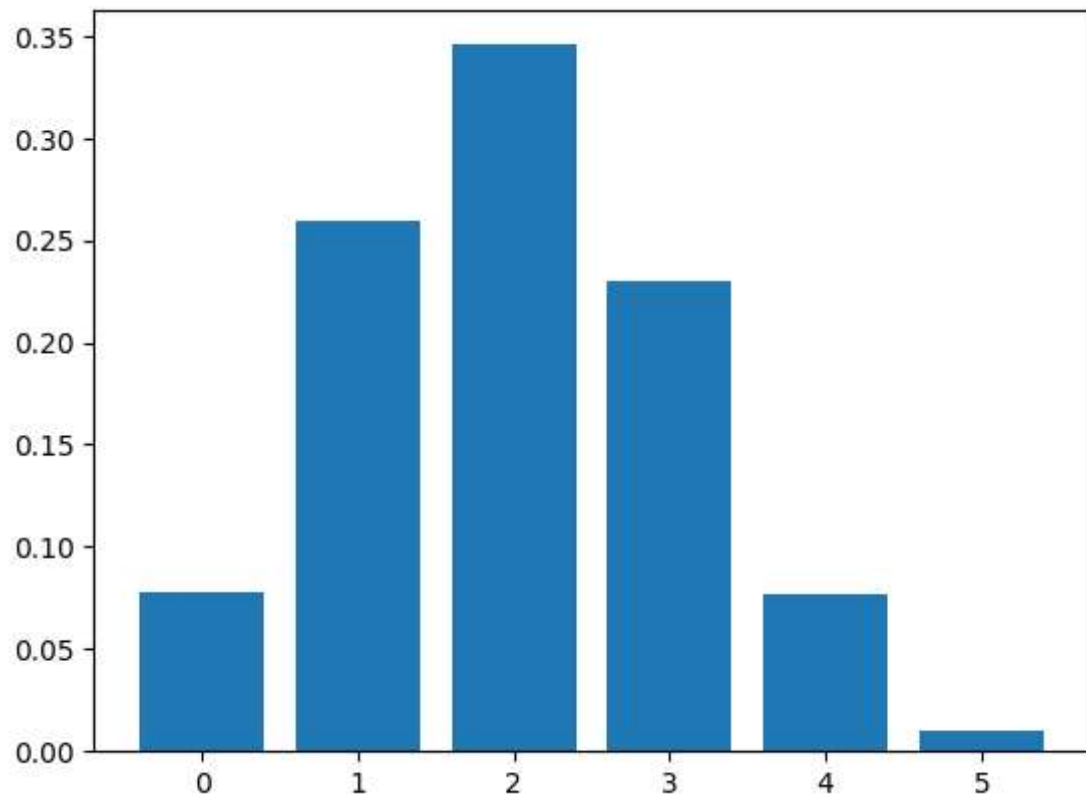
In [27]: 1 pip install matplotlib

Requirement already satisfied: matplotlib in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (3.7.1)  
Requirement already satisfied: contourpy>=1.0.1 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (1.0.7)  
Requirement already satisfied: cyclor>=0.10 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (0.11.0)  
Requirement already satisfied: fonttools>=4.22.0 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (4.39.4)  
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (1.4.4)  
Requirement already satisfied: numpy>=1.20 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (1.24.3)  
Requirement already satisfied: packaging>=20.0 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (23.1)  
Requirement already satisfied: pillow>=6.2.0 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (9.5.0)  
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (3.0.9)  
Requirement already satisfied: python-dateutil>=2.7 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (2.8.2)  
Requirement already satisfied: six>=1.5 in c:\users\teppa\appdata\local\programs\python\python310\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)  
Note: you may need to restart the kernel to use updated packages.

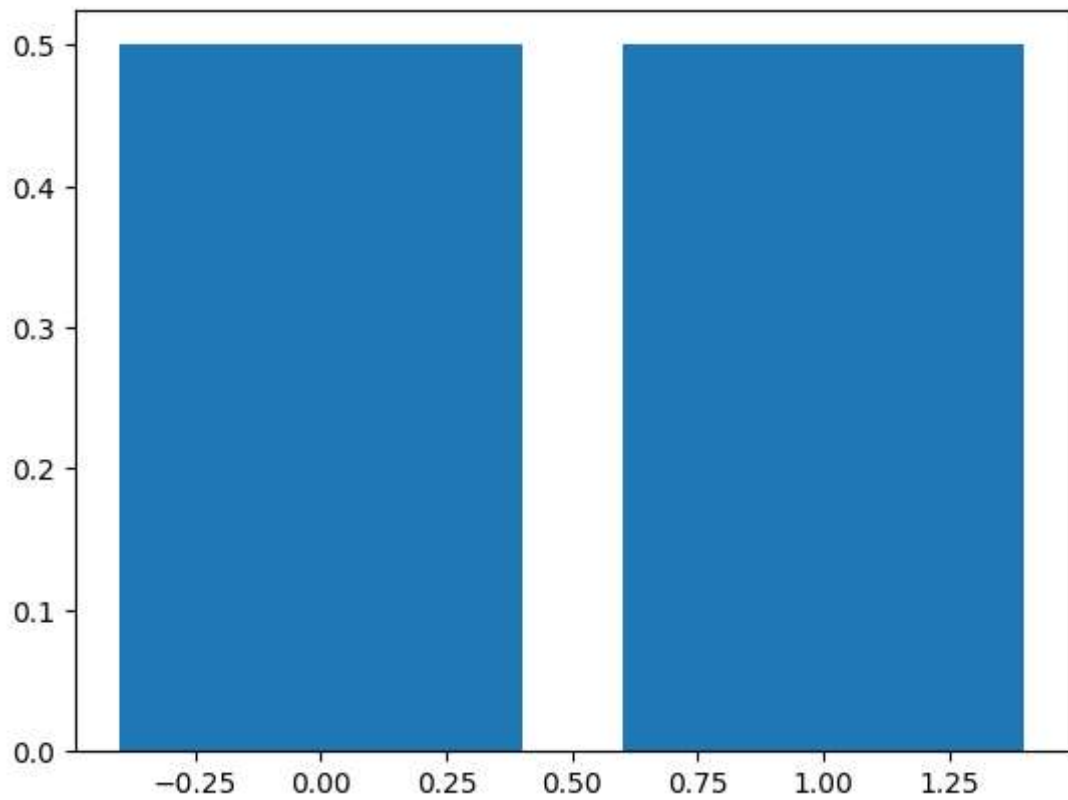
In [28]: 1 import matplotlib.pyplot as plt

In [29]:

```
1 n=5
2 p=0.4
3 r_values=list(range(n+1))
4 dist=[binom.pmf(r,n,p) for r in r_values]
5 plt.bar(r_values,dist)
6 plt.show()
```

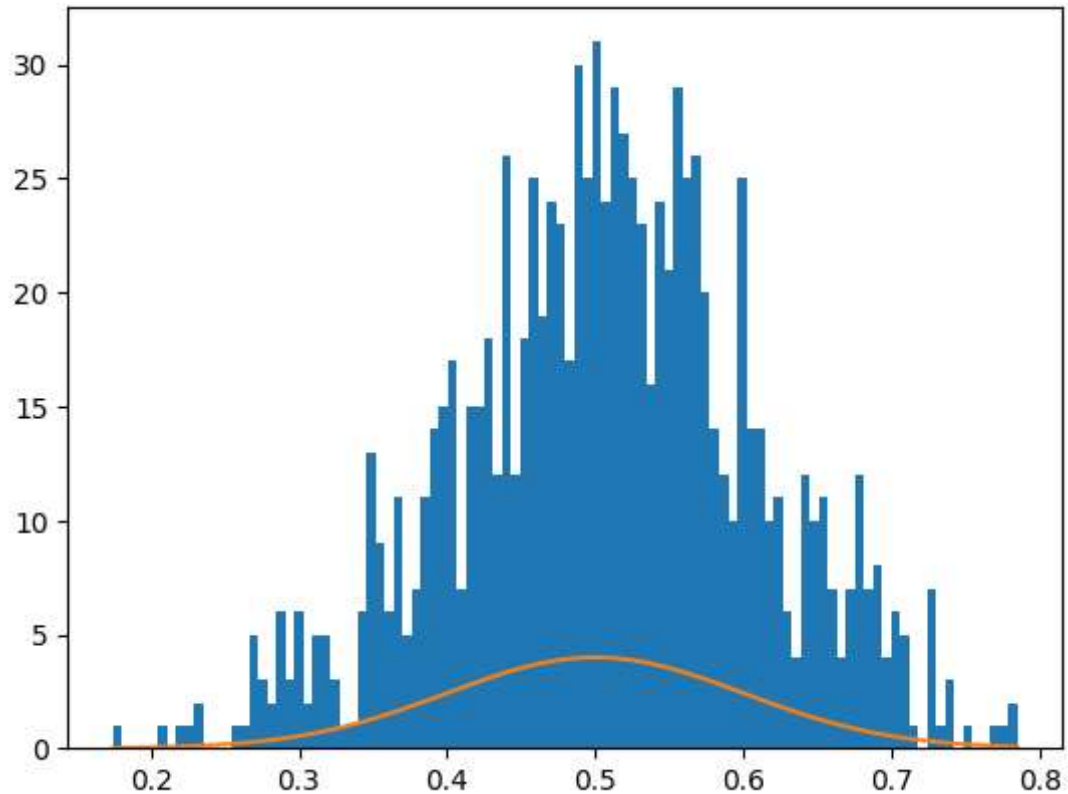


```
In [30]: 1 import matplotlib.pyplot as plt
2 from scipy.stats import bernoulli
3 bd=bernoulli(0.5)
4 X=[0,1]
5 plt.bar(X,bd.pmf(X))
6 plt.show()
```



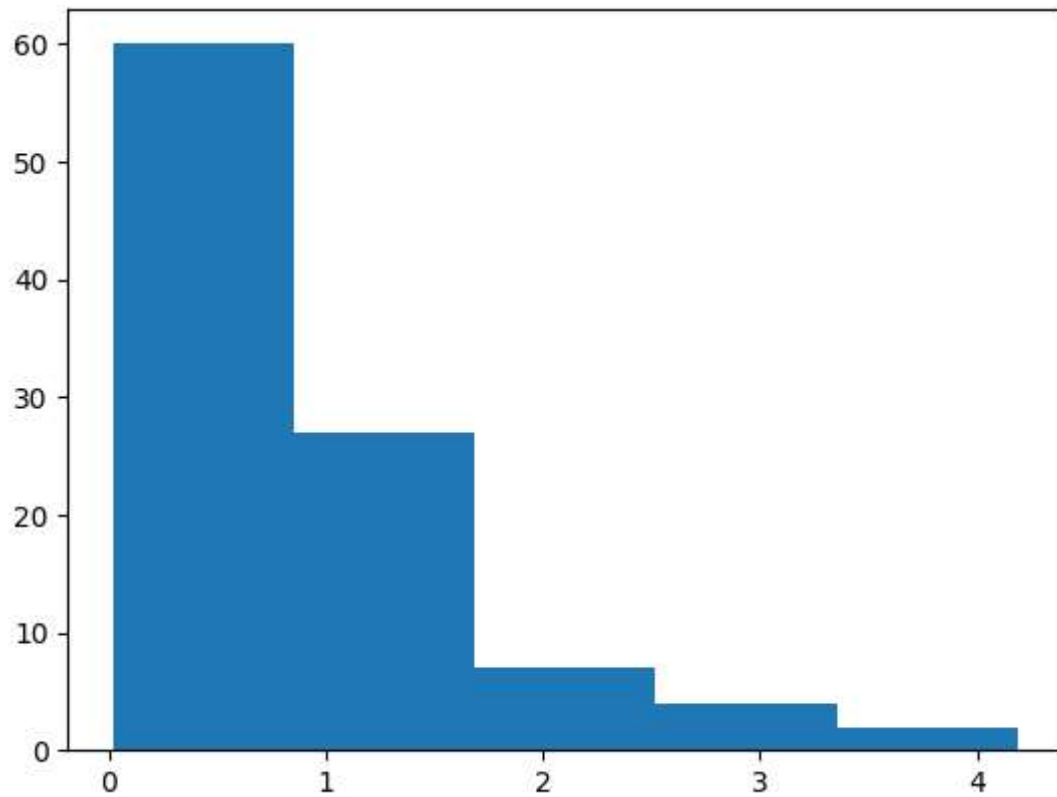
```
In [ ]: 1
```

```
In [31]: 1 import matplotlib.pyplot as plt
2 import numpy as np
3 mu,sigma=0.5,0.1
4 s=np.random.normal(mu,sigma,1000)
5 count,bins,ignored=plt.hist(s,100)
6 plt.plot(bins,1/(sigma*np.sqrt(2*np.pi))*np.exp(-(bins-mu)**2/(2*sigma**2)
7 plt.show()
```





```
In [32]: 1 import numpy as np
          2 import matplotlib.pyplot as plt
          3 exp=np.random.exponential(1,100)
          4 count,bins,ignored=plt.hist(exp,5)
          5 plt.show()
```



```
In [ ]: 1
```