```
import numpy as np
arr=np.array([2,3,4,6,7,8,9,12,13,16,17,23,25,27,34,37,201])

arr
```

```
array([  2,   3,   4,   6,   7,   8,   9,  12,  13,  16,  17,  23,
25,
        27,  34,  37, 201])
```

```
q1=np.percentile(arr,25)
q3=np.percentile(arr,75)
IQR=q3-q1
first=q1-(1.5*IQR)
last=q3+(1.5*IQR)
print(q1)
print(q3)
print(IQR)
print(first)
print(last)
```

```
7.0
25.0
18.0
-20.0
52.0
```

```
arr=arr[(arr>first) & (arr<last)]
arr
```

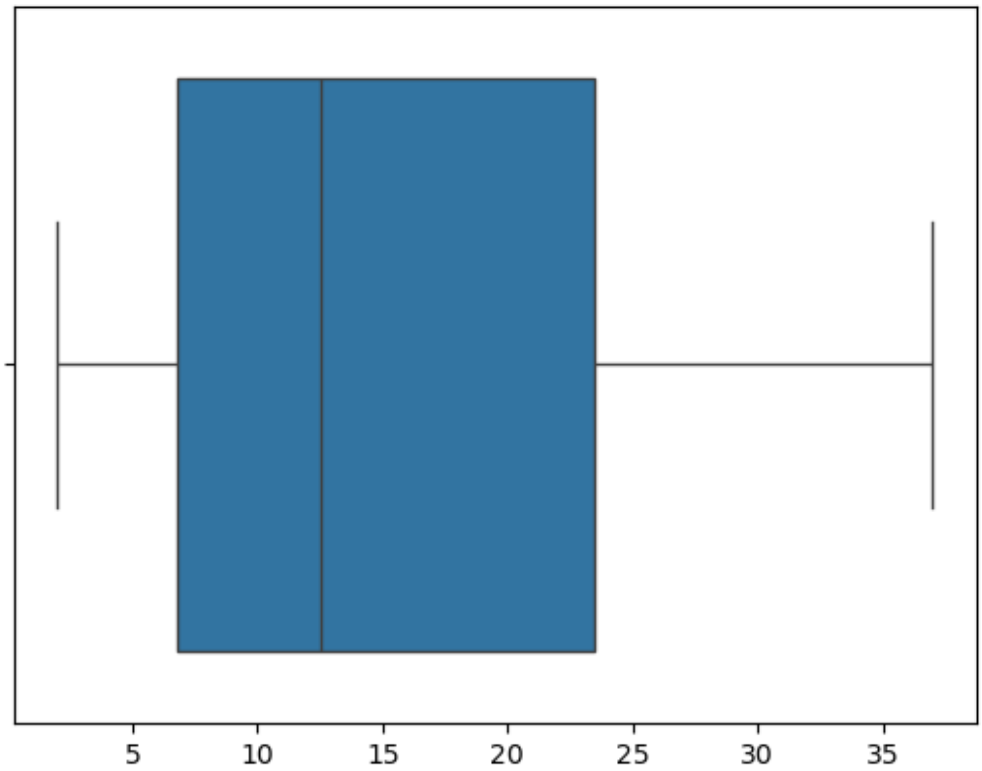
```
array([ 2,  3,  4,  6,  7,  8,  9, 12, 13, 16, 17, 23, 25, 27, 34,
37])
```

```
max=np.max(arr)
min=np.min(arr)
q1=np.percentile(arr,25)
q3=np.percentile(arr,75)
mid=np.median(arr)
print(min)
print(q1)
print(mid)
print(q3)
print(max)
```

```
2
6.75
12.5
23.5
37
```

```
import seaborn as sns
sns.boxplot(x=arr)
```

```
<Axes: >
```



Z TEST

```
from scipy.stats import norm

sample=[172,174,168,169,171,173,175,170,169,172]

popluation_mean=170
population_std=3
samp_mean=np.mean(sample)
n=len(sample)

z_score=(samp_mean-popluation_mean)/(population_std/np.sqrt(n))
z_score
```

```
np.float64(1.3703203194063098)
```

```
p_value=2*(1-norm.cdf(z_score))  #formula

p_value
```

```
np.float64(0.17058693287143756)
```

```
alpha=0.05
```

result

- if aplha>=p_value => we reject null hypothesis
- if alpha<p_value=> we accept null hypothesis

```
if p_value<=alpha:
    print("Reject null hypthesis")
elif p_value>alpha:
    print("Accept null hypthesis")
```

```
Accept null hypthesis
```

T test

```
from scipy import stats
sample=[172,174,168,169,171,173,175,170,169,172]

population_mean=170
sample_mean=np.mean(sample)
sample_std=np.std(sample,ddof=1)
n=len(sample)

t_stat=(sample_mean-popluation_mean)/(sample_std/np.sqrt(n))
t_stat
```

```
np.float64(1.7782469350914734)
```

```
p_value=2*(1-stats.t.cdf(abs(t_stat),df=n-1))
p_value
```

```
np.float64(0.10907771593031335)
```

```
alpha=0.05

if p_value<=alpha:
    print("Reject null hypthesis")
elif p_value>alpha:
    print("Accept null hypthesis")
```

```
Accept null hypthesis
```

same as z test

result

- if aplha>=p_value => we reject null hypothesis
- if alpha<p_value=> we accept null hypothesis

TWO SAMPLE TEST

```
group_A=[85,88,90,92,87,85,89,91,86,88]
group_B=[82,84,80,83,81,79,78,85,84,83]

t_stat,p_value=stats.ttest_ind(group_A,group_B,equal_var=False)

t_stat

np.float64(5.829604009507161)

p_value

np.float64(1.610475598965881e-05)

alpha=0.05

if p_value<=alpha:
    print("Reject null hypthesis")
elif p_value>alpha:
    print("Accept null hypthesis")

Reject null hypthesis
```

CHI SQUARE TEST

```
import pandas as pd
df = sns.load_dataset('titanic')
from scipy.stats import chi2_contingency

df.head(3)

   survived  pclass     sex   age  sibsp  parch     fare embarked
class  \
0         0       3    male  22.0      1      0   7.2500        S
Third
1         1       1  female  38.0      1      0  71.2833        C
First
2         1       3  female  26.0      0      0   7.9250        S
Third

     who  adult_male deck  embark_town alive  alone
0    man        True  NaN  Southampton    no  False
1  woman       False    C    Cherbourg   yes  False
2  woman       False  NaN  Southampton   yes   True

contigency_table=pd.crosstab(df['sex'],df['survived'])

contigency_table

survived    0    1
sex
```

```
female     81  233
male      468  109

chi2,p_value,dof,expected=chi2_contingency(contigency_table)

chi2

np.float64(260.71702016732104)

p_value

np.float64(1.1973570627755645e-58)

dof

1

expected

array([[193.47474747, 120.52525253],
       [355.52525253, 221.47474747]])

if p_value<=alpha:
    print("there is a significant relationship btw gender and
survival")
elif p_value>alpha:
    print("Accept null hypthesis")

there is a significant relationship btw gender and survival
```

ANNOVA TEST

```
from scipy.stats import f_oneway

df.head(3)

   survived  pclass     sex   age  sibsp  parch     fare embarked
class  \
0         0       3    male  22.0      1      0   7.2500        S
Third
1         1       1  female  38.0      1      0  71.2833        C
First
2         1       3  female  26.0      0      0   7.9250        S
Third

     who  adult_male deck  embark_town alive  alone
0    man        True  NaN  Southampton    no  False
1  woman       False    C    Cherbourg   yes  False
2  woman       False  NaN  Southampton   yes   True

df=df[['age','pclass']].dropna()
```

```
class1=df[df['pclass'] == 1]['age']

class2=df[df['pclass'] == 2]['age']

class3=df[df['pclass'] == 3]['age']

class1
```

```
1      38.0
3      35.0
6      54.0
11     58.0
23     28.0
       ...
871    47.0
872    33.0
879    56.0
887    19.0
889    26.0
Name: age, Length: 186, dtype: float64
```

```
class2
```

```
9      14.0
15     55.0
20     35.0
21     34.0
33     66.0
       ...
866    27.0
874    28.0
880    25.0
883    28.0
886    27.0
Name: age, Length: 173, dtype: float64
```

```
class3
```

```
0      22.0
2      26.0
4      35.0
7       2.0
8      27.0
       ...
881    33.0
882    22.0
884    25.0
885    39.0
890    32.0
Name: age, Length: 355, dtype: float64
```

```
f_stats,p_value=f_oneway(class1,class2,class3)
```

```
f_stats
```

```
np.float64(57.443484340676214)
```

```
p_value
```

```
np.float64(7.487984171959904e-24)
```

```
alpha=0.05
```

```
if p_value<=alpha:
    print("there is a significant difference btw age group and
passenger class ")
elif p_value>alpha:
    print("Accept null hypthesis")
```

```
there is a significant difference btw age group and passenger class
```