dav-zscore

March 22, 2024

Zscore for outlier detection #Zscore=(x-mean)/std.deviation

```
[12]: import numpy as np
      data=[1,2,3,2,2,3,1,1,1,1,15,23,2,3,3,3,1]
      mean=np.mean(data)
      std=np.std(data)
      print("mean of the dataset is ",mean)
      print("std.devition is",std)
      threshold=3
      outlier=[]
      for i in data:
        z=(i-mean)/std
        if z>threshold:
          outlier.append(i)
      print(outlier)
```

```
mean of the dataset is  3.9411764705882355
std.devition is 5.72374402596242
[23]
```

#IQR=Q3-Q1 Q1 represents the 25th percentile.

Q2 represents the 50th percentile.

Q3 represents the 75th percentile.

Q1=median of the dataset

Q2=median of n smallest data points

Q3=median of n highesh data points

```
[13]: import numpy as np
      import seaborn as sns
      data=[6,4,5,30,2,8,16]
      sort_data=np.sort(data)
      sort_data
```

```
[13]: array([ 2,  4,  5,  6,  8, 16, 30])
```

```
[14]: Q1=np.percentile(data,25,interpolation='midpoint')
      Q2=np.percentile(data,50,interpolation='midpoint')
      Q3=np.percentile(data,75,interpolation='midpoint')

      print('Q1 25 percentile of the given data is',Q1)
      print('Q2 50 percentile of the given data is',Q2)
      print('Q3 75 percentile of the given data is',Q3)

      IQR=Q3-Q1
      print('interquartile range is',IQR)
```

```
Q1 25 percentile of the given data is 4.5
Q2 50 percentile of the given data is 6.0
Q3 75 percentile of the given data is 12.0
interquartile range is 7.5
```

```
[16]: low_lim=Q1-1.5*IQR
      up_lim=Q3+1.5*IQR
      print(low_lim)
      print(up_lim)
```
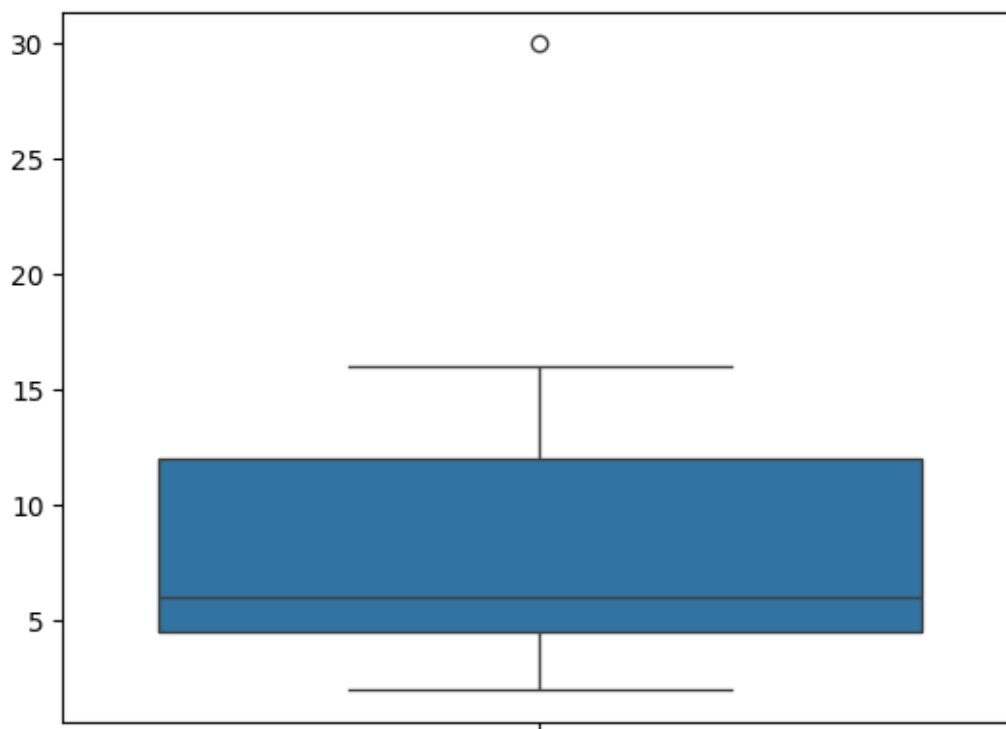
```
-6.75
23.25
```

```
[17]: outlier=[]
      for x in data:
        if((x>up_lim)or(x<low_lim)):
          outlier.append(x)
      print('outlier',outlier)
```

```
outlier [30]
```

```
[18]: sns.boxplot(data)
```

```
[18]: <Axes: >
```