

In [102]:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import skew
import seaborn as sns
```

In [103]:

```
df=pd.read_csv('/home/anaconda/Downloads/Dataset/StudentsPerformance.csv')
```

In [104]:

```
df1 =df
```

In [105]:

```
df.describe()
```

Out[105]:

	math score	reading score	writing score
count	987.000000	975.000000	997.000000
mean	66.310030	69.007179	68.020060
std	14.954533	14.557997	15.202634
min	8.000000	17.000000	10.000000
25%	57.000000	59.000000	57.000000
50%	66.000000	70.000000	69.000000
75%	77.000000	79.000000	79.000000
max	100.000000	100.000000	100.000000

In [106]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
gender                1000 non-null object
race/ethnicity        1000 non-null object
parental level of education  1000 non-null object
lunch                 1000 non-null object
test preparation course  1000 non-null object
math score            987 non-null float64
reading score         975 non-null float64
writing score         997 non-null float64
dtypes: float64(3), object(5)
memory usage: 62.6+ KB
```

In [107]:

```
df.dtypes
```

Out[107]:

```
gender                object
race/ethnicity        object
parental level of education  object
lunch                 object
test preparation course  object
math score            float64
reading score         float64
writing score         float64
```

```
writing score          float64
dtype: object
```

```
In [108]:
```

```
df.isnull().sum()
```

```
Out[108]:
```

```
gender                0
race/ethnicity        0
parental level of education  0
lunch                 0
test preparation course  0
math score            13
reading score         25
writing score         3
dtype: int64
```

```
In [109]:
```

```
df.shape
```

```
Out[109]:
```

```
(1000, 8)
```

```
In [110]:
```

```
df1.isnull().sum()
```

```
Out[110]:
```

```
gender                0
race/ethnicity        0
parental level of education  0
lunch                 0
test preparation course  0
math score            13
reading score         25
writing score         3
dtype: int64
```

```
In [111]:
```

```
df1=df1.interpolate()
df1.isnull().sum()
```

```
Out[111]:
```

```
gender                0
race/ethnicity        0
parental level of education  0
lunch                 0
test preparation course  0
math score            0
reading score         0
writing score         0
dtype: int64
```

```
In [112]:
```

```
q1=df1['math score'].quantile(0.25)
q3=df1['math score'].quantile(0.75)
q1,q3
```

```
Out[112]:
```

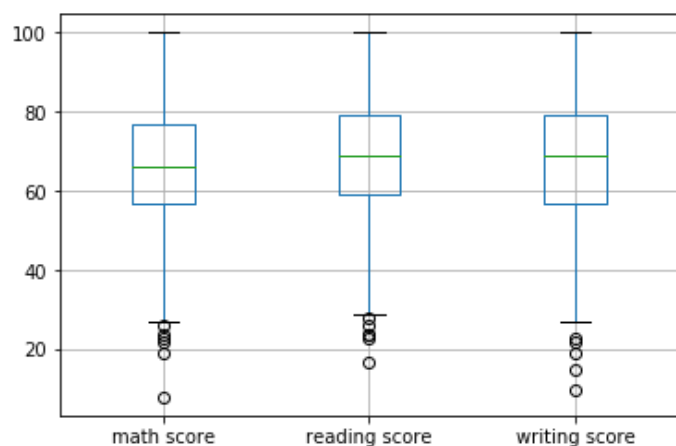
```
(57.0, 77.0)
```

```
In [113]:
```

```
df1.boxplot()
```

Out[113]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f49a2d2b208>



In [114]:

```
iqr=q3-q1  
iqr
```

Out[114]:

20.0

In [115]:

```
upper=q1-(1.5*iqr)  
lower=q3+(1.5*iqr)  
upper , lower
```

Out[115]:

(27.0, 107.0)

In [116]:

```
final1=df1[df1['math score']< lower]  
final2=df1[df1['math score']>upper]  
  
final1
```

Out[116]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72.000000	72.0	74.0
1	female	group C	some college	standard	completed	69.000000	90.0	88.0
2	female	group B	master's degree	standard	none	90.000000	95.0	93.0
3	male	group A	associate's degree	free/reduced	none	83.000000	57.0	44.0
4	male	group C	some college	standard	none	76.000000	70.0	75.0
5	female	group B	associate's degree	standard	none	71.000000	83.0	78.0
6	female	group B	some college	standard	completed	88.000000	95.0	92.0
7	male	group B	some college	free/reduced	none	40.000000	43.0	39.0
8	male	group D	high school	free/reduced	completed	64.000000	64.0	67.0
9	female	group B	high school	free/reduced	none	38.000000	60.0	50.0
10	male	group C	associate's degree	standard	none	58.000000	54.0	52.0
11	male	group D	associate's degree	standard	none	40.000000	52.0	43.0
12	female	group B	high school	standard	none	65.000000	81.0	73.0
13	male	group A	some college	standard	completed	78.000000	72.0	70.0

14	female	group A	parental level of education	standard lunch	test preparation course	50.000000	reading score	writing score
15	female	group C	some high school	standard	none	60.000000	75.0	78.0
16	male	group C	high school	standard	none	88.000000	89.0	53.0
17	female	group B	some high school	free/reduced	none	76.666667	32.0	28.0
18	male	group C	master's degree	free/reduced	completed	65.333333	42.0	46.0
19	female	group C	associate's degree	free/reduced	none	54.000000	58.0	61.0
20	male	group D	high school	standard	none	66.000000	69.0	63.0
21	female	group B	some college	free/reduced	completed	65.000000	75.0	70.0
22	male	group D	some college	standard	none	44.000000	54.0	53.0
23	female	group C	some high school	standard	none	69.000000	73.0	73.0
24	male	group D	bachelor's degree	free/reduced	completed	74.000000	71.0	67.0
25	male	group A	master's degree	free/reduced	none	73.000000	74.0	61.0
26	male	group B	some college	standard	none	69.000000	54.0	55.0
27	female	group C	bachelor's degree	standard	none	67.000000	69.0	75.0
28	male	group C	high school	standard	none	70.000000	70.0	65.0
29	female	group D	master's degree	standard	none	62.000000	70.0	75.0
...
970	female	group D	bachelor's degree	standard	none	89.000000	100.0	100.0
971	male	group C	some high school	standard	completed	78.000000	72.0	69.0
972	female	group A	high school	free/reduced	completed	53.000000	50.0	60.0
973	female	group D	some college	free/reduced	none	49.000000	65.0	61.0
974	female	group A	some college	standard	none	54.000000	63.0	67.0
975	female	group C	some college	standard	completed	64.000000	82.0	77.0
976	male	group B	some college	free/reduced	completed	60.000000	62.0	60.0
977	male	group C	associate's degree	standard	none	62.000000	65.0	58.0
978	male	group D	high school	standard	completed	55.000000	41.0	48.0
979	female	group C	associate's degree	standard	none	91.000000	95.0	94.0
980	female	group B	high school	free/reduced	none	8.000000	24.0	23.0
981	male	group D	some high school	standard	none	81.000000	78.0	78.0
982	male	group B	some high school	standard	completed	79.000000	85.0	86.0
983	female	group A	some college	standard	completed	78.000000	87.0	91.0
984	female	group C	some high school	standard	none	74.000000	75.0	82.0
985	male	group A	high school	standard	none	57.000000	51.0	54.0
986	female	group C	associate's degree	standard	none	40.000000	59.0	51.0
987	male	group E	some high school	standard	completed	81.000000	75.0	76.0
988	female	group A	some high school	free/reduced	none	44.000000	45.0	45.0
989	female	group D	some college	free/reduced	completed	67.000000	86.0	83.0
990	male	group E	high school	free/reduced	completed	86.000000	81.0	75.0
991	female	group B	some high school	standard	completed	65.000000	82.0	78.0
992	female	group D	associate's degree	free/reduced	none	55.000000	76.0	76.0
993	female	group D	bachelor's degree	free/reduced	none	62.000000	72.0	74.0
994	male	group A	high school	standard	none	63.000000	63.0	62.0
995	female	group E	master's degree	standard	completed	88.000000	99.0	95.0
996	male	group C	high school	free/reduced	none	62.000000	55.0	55.0
997	female	group C	high school	free/reduced	completed	59.000000	71.0	65.0

998	female	group D	bachelor's degree	standard	test preparation course	68.000000	reading score	writing score
999	female	group D	some college	free/reduced	none	77.000000	86.0	86.0

1000 rows x 8 columns

In [124]:

```
x=np.random.normal(0,2,1000)
skew(x)
```

Out[124]:

0.031669794546370415

In [125]:

```
final2=df1[df1['math score']>upper]
final2
```

Out[125]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72.000000	72.0	74.0
1	female	group C	some college	standard	completed	69.000000	90.0	88.0
2	female	group B	master's degree	standard	none	90.000000	95.0	93.0
3	male	group A	associate's degree	free/reduced	none	83.000000	57.0	44.0
4	male	group C	some college	standard	none	76.000000	70.0	75.0
5	female	group B	associate's degree	standard	none	71.000000	83.0	78.0
6	female	group B	some college	standard	completed	88.000000	95.0	92.0
7	male	group B	some college	free/reduced	none	40.000000	43.0	39.0
8	male	group D	high school	free/reduced	completed	64.000000	64.0	67.0
9	female	group B	high school	free/reduced	none	38.000000	60.0	50.0
10	male	group C	associate's degree	standard	none	58.000000	54.0	52.0
11	male	group D	associate's degree	standard	none	40.000000	52.0	43.0
12	female	group B	high school	standard	none	65.000000	81.0	73.0
13	male	group A	some college	standard	completed	78.000000	72.0	70.0
14	female	group A	master's degree	standard	none	50.000000	53.0	58.0
15	female	group C	some high school	standard	none	69.000000	75.0	78.0
16	male	group C	high school	standard	none	88.000000	89.0	53.0
17	female	group B	some high school	free/reduced	none	76.666667	32.0	28.0
18	male	group C	master's degree	free/reduced	completed	65.333333	42.0	46.0
19	female	group C	associate's degree	free/reduced	none	54.000000	58.0	61.0
20	male	group D	high school	standard	none	66.000000	69.0	63.0
21	female	group B	some college	free/reduced	completed	65.000000	75.0	70.0
22	male	group D	some college	standard	none	44.000000	54.0	53.0
23	female	group C	some high school	standard	none	69.000000	73.0	73.0
24	male	group D	bachelor's degree	free/reduced	completed	74.000000	71.0	67.0
25	male	group A	master's degree	free/reduced	none	73.000000	74.0	61.0
26	male	group B	some college	standard	none	69.000000	54.0	55.0
27	female	group C	bachelor's degree	standard	none	67.000000	69.0	75.0
28	male	group C	high school	standard	none	70.000000	70.0	65.0
29	female	group D	master's degree	standard	none	62.000000	70.0	75.0

...	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
969	female	group B	bachelor's degree	standard	none	75.000000	84.0	80.0
970	female	group D	bachelor's degree	standard	none	89.000000	100.0	100.0
971	male	group C	some high school	standard	completed	78.000000	72.0	69.0
972	female	group A	high school	free/reduced	completed	53.000000	50.0	60.0
973	female	group D	some college	free/reduced	none	49.000000	65.0	61.0
974	female	group A	some college	standard	none	54.000000	63.0	67.0
975	female	group C	some college	standard	completed	64.000000	82.0	77.0
976	male	group B	some college	free/reduced	completed	60.000000	62.0	60.0
977	male	group C	associate's degree	standard	none	62.000000	65.0	58.0
978	male	group D	high school	standard	completed	55.000000	41.0	48.0
979	female	group C	associate's degree	standard	none	91.000000	95.0	94.0
981	male	group D	some high school	standard	none	81.000000	78.0	78.0
982	male	group B	some high school	standard	completed	79.000000	85.0	86.0
983	female	group A	some college	standard	completed	78.000000	87.0	91.0
984	female	group C	some high school	standard	none	74.000000	75.0	82.0
985	male	group A	high school	standard	none	57.000000	51.0	54.0
986	female	group C	associate's degree	standard	none	40.000000	59.0	51.0
987	male	group E	some high school	standard	completed	81.000000	75.0	76.0
988	female	group A	some high school	free/reduced	none	44.000000	45.0	45.0
989	female	group D	some college	free/reduced	completed	67.000000	86.0	83.0
990	male	group E	high school	free/reduced	completed	86.000000	81.0	75.0
991	female	group B	some high school	standard	completed	65.000000	82.0	78.0
992	female	group D	associate's degree	free/reduced	none	55.000000	76.0	76.0
993	female	group D	bachelor's degree	free/reduced	none	62.000000	72.0	74.0
994	male	group A	high school	standard	none	63.000000	63.0	62.0
995	female	group E	master's degree	standard	completed	88.000000	99.0	95.0
996	male	group C	high school	free/reduced	none	62.000000	55.0	55.0
997	female	group C	high school	free/reduced	completed	59.000000	71.0	65.0
998	female	group D	some college	standard	completed	68.000000	78.0	77.0
999	female	group D	some college	free/reduced	none	77.000000	86.0	86.0

992 rows x 8 columns

In [129]:

```
final2["math score"].shape
```

Out[129]:

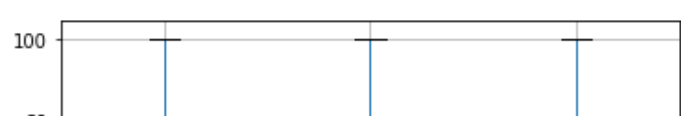
(992,)

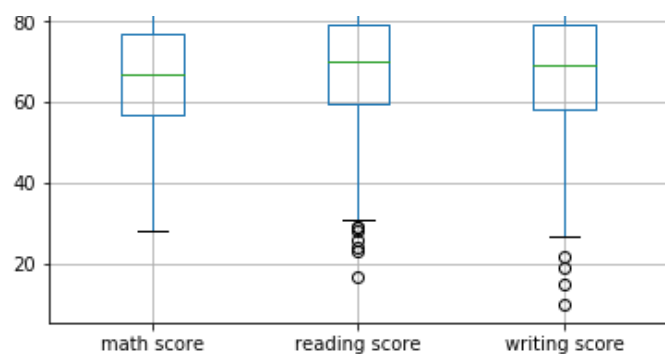
In [131]:

```
final2.boxplot()
```

Out[131]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f49a2a43e80>





In [120]:

```
x=np.random.normal(0,2,1000)
```

In [121]:

```
skew(x)
```

Out[121]:

```
-0.06686610417413578
```

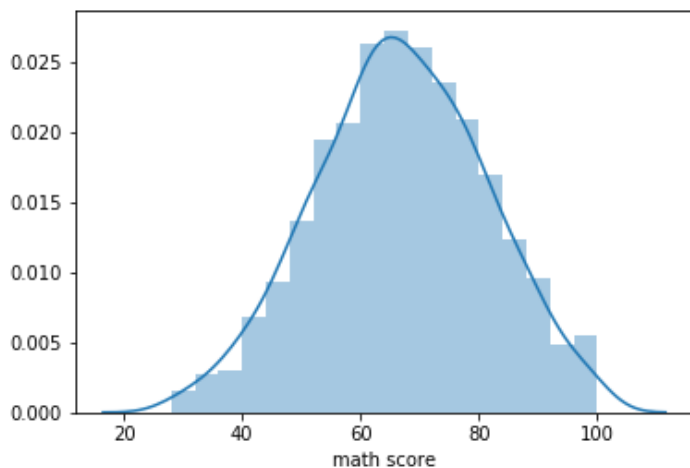
In [128]:

```
sns.distplot(final2['math score'])
```

/home/anaconda/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval

Out[128]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f49a2ac04e0>
```



In []:

In []: