In [22]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy import stats
```

In [30]:
```
data = {
    'StudentId': range(1, 31),
    'Name': [
        'Abhishek Podi','Abhishek Bachhan','Bikash Pandey','Anish Joshi','Abhay Sharm
        'Ashish Song','Ajay Cinthol','Ashutosh Song','Ayush Roll17','Abhishek Poda',
        'Rohit Kumar','Neha Singh','Pooja Patil','Suresh Mehta','Kunal Shah',
        'Aman Verma','Riya Desai','Vikas Rao','Sneha Nair','Mohit Jain',
        'Kriti Malhotra','Arjun Singh','Priya Kapoor','Nitin Yadav','Shreya Bose',
        'Deepak Gupta','Simran Kaur','Rahul Mishra','Tina Roy','Manish Tiwari'
    ],
    'Mathematics': [
        98, 85, 76, 23, np.nan, 100, 125, 55, 21, 88,
        92, 81, 77, 66, 59, 101, 34, 48, 72, 83,
        150, 95, 67, 29, 74, -20, 90, 86, 60, 79
    ],
    'Physics': [
        78, 55, 96, 150, 46, 66, np.nan, 35, 81, 78,
        88, 72, 69, 58, 61, 120, 42, 39, 75, 84,
        160, 91, 70, 28, 73, -10, 89, 87, 62, 80
    ],
    'Chemistry': [
        -95, 45, np.nan, 73, 76, 79, 86, 65, 41, np.nan,
        88, 74, 71, 60, 63, 110, 38, 44, 70, 82,
        145, 90, 68, 25, 72, -30, 91, 85, 59, 77
    ]
}

df = pd.DataFrame(data)
df
```

Out[30]:

	StudentId	Name	Mathematics	Physics	Chemistry
0	1	Abhishek Podi	98.0	78.0	-95.0
1	2	Abhishek Bachhan	85.0	55.0	45.0
2	3	Bikash Pandey	76.0	96.0	NaN
3	4	Anish Joshi	23.0	150.0	73.0
4	5	Abhay Sharma	NaN	46.0	76.0
5	6	Ashish Song	100.0	66.0	79.0
6	7	Ajay Cinthol	125.0	NaN	86.0
7	8	Ashutosh Song	55.0	35.0	65.0
8	9	Ayush Roll17	21.0	81.0	41.0
9	10	Abhishek Poda	88.0	78.0	NaN
10	11	Rohit Kumar	92.0	88.0	88.0
11	12	Neha Singh	81.0	72.0	74.0
12	13	Pooja Patil	77.0	69.0	71.0
13	14	Suresh Mehta	66.0	58.0	60.0
14	15	Kunal Shah	59.0	61.0	63.0
15	16	Aman Verma	101.0	120.0	110.0
16	17	Riya Desai	34.0	42.0	38.0
17	18	Vikas Rao	48.0	39.0	44.0
18	19	Sneha Nair	72.0	75.0	70.0
19	20	Mohit Jain	83.0	84.0	82.0
20	21	Kriti Malhotra	150.0	160.0	145.0
21	22	Arjun Singh	95.0	91.0	90.0
22	23	Priya Kapoor	67.0	70.0	68.0
23	24	Nitin Yadav	29.0	28.0	25.0
24	25	Shreya Bose	74.0	73.0	72.0
25	26	Deepak Gupta	-20.0	-10.0	-30.0
26	27	Simran Kaur	90.0	89.0	91.0
27	28	Rahul Mishra	86.0	87.0	85.0
28	29	Tina Roy	60.0	62.0	59.0
29	30	Manish Tiwari	79.0	80.0	77.0

In [32]:
```
df.isnull().sum()
```

Out[32]:
```
StudentId      0
Name           0
Mathematics    1
Physics        1
Chemistry      2
dtype: int64
```

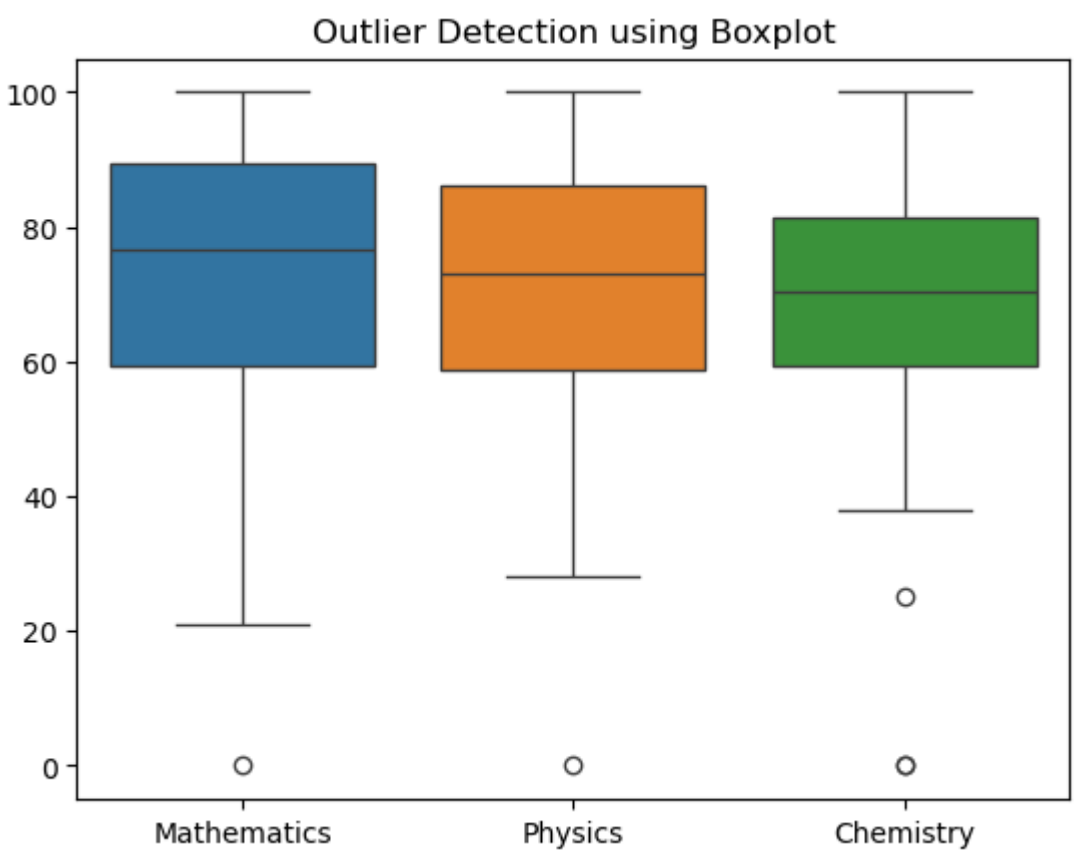
In [34]:
```
col = ['Mathematics', 'Physics', 'Chemistry']
df[col] = df[col].fillna(df[col].mean())
df
```

Out[34]:

	StudentId	Name	Mathematics	Physics	Chemistry
0	1	Abhishek Podi	98.000000	78.000000	-95.000000
1	2	Abhishek Bachhan	85.000000	55.000000	45.000000
2	3	Bikash Pandey	76.000000	96.000000	62.571429
3	4	Anish Joshi	23.000000	150.000000	73.000000
4	5	Abhay Sharma	72.206897	46.000000	76.000000
5	6	Ashish Song	100.000000	66.000000	79.000000
6	7	Ajay Cinthol	125.000000	73.206897	86.000000
7	8	Ashutosh Song	55.000000	35.000000	65.000000
8	9	Ayush Roll17	21.000000	81.000000	41.000000
9	10	Abhishek Poda	88.000000	78.000000	62.571429
10	11	Rohit Kumar	92.000000	88.000000	88.000000
11	12	Neha Singh	81.000000	72.000000	74.000000
12	13	Pooja Patil	77.000000	69.000000	71.000000
13	14	Suresh Mehta	66.000000	58.000000	60.000000
14	15	Kunal Shah	59.000000	61.000000	63.000000
15	16	Aman Verma	101.000000	120.000000	110.000000
16	17	Riya Desai	34.000000	42.000000	38.000000
17	18	Vikas Rao	48.000000	39.000000	44.000000
18	19	Sneha Nair	72.000000	75.000000	70.000000
19	20	Mohit Jain	83.000000	84.000000	82.000000
20	21	Kriti Malhotra	150.000000	160.000000	145.000000
21	22	Arjun Singh	95.000000	91.000000	90.000000
22	23	Priya Kapoor	67.000000	70.000000	68.000000
23	24	Nitin Yadav	29.000000	28.000000	25.000000
24	25	Shreya Bose	74.000000	73.000000	72.000000
25	26	Deepak Gupta	-20.000000	-10.000000	-30.000000
26	27	Simran Kaur	90.000000	89.000000	91.000000
27	28	Rahul Mishra	86.000000	87.000000	85.000000
28	29	Tina Roy	60.000000	62.000000	59.000000
29	30	Manish Tiwari	79.000000	80.000000	77.000000

In [36]:
```
for col in ['Mathematics', 'Physics', 'Chemistry']:
    df.loc[df[col] > 100, col] = 100
    df.loc[df[col] < 0, col] = 0
```

In [38]:
```
sns.boxplot(data=df[['Mathematics', 'Physics', 'Chemistry']])
plt.title("Outlier Detection using Boxplot")
plt.show()
```



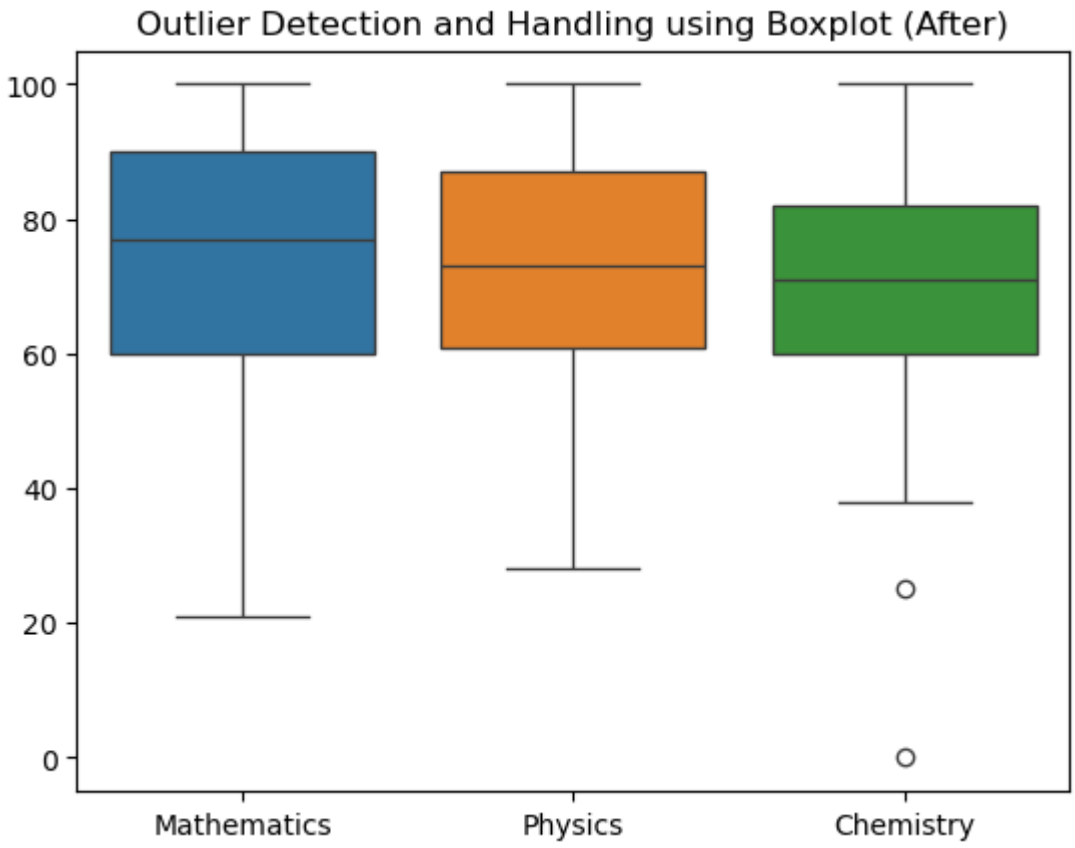
In [40]:
```
Q1 = df['Mathematics'].quantile(0.25)
Q3 = df['Mathematics'].quantile(0.75)
IQR = Q3 - Q1

lower = Q1 - 1.5 * IQR
upper = Q3 + 1.5 * IQR

df['Mathematics'] = np.where(df['Mathematics'] > upper, upper, np.where(df['Mathemati
```

In [71]:
```
z_scores = np.abs(stats.zscore(df['Physics']))
df = df[z_scores < 2.5]
```

In [73]:
```
sns.boxplot(data=df[['Mathematics','Physics','Chemistry']])
plt.title("Outlier Detection and Handling using Boxplot (After)")
plt.show()
```



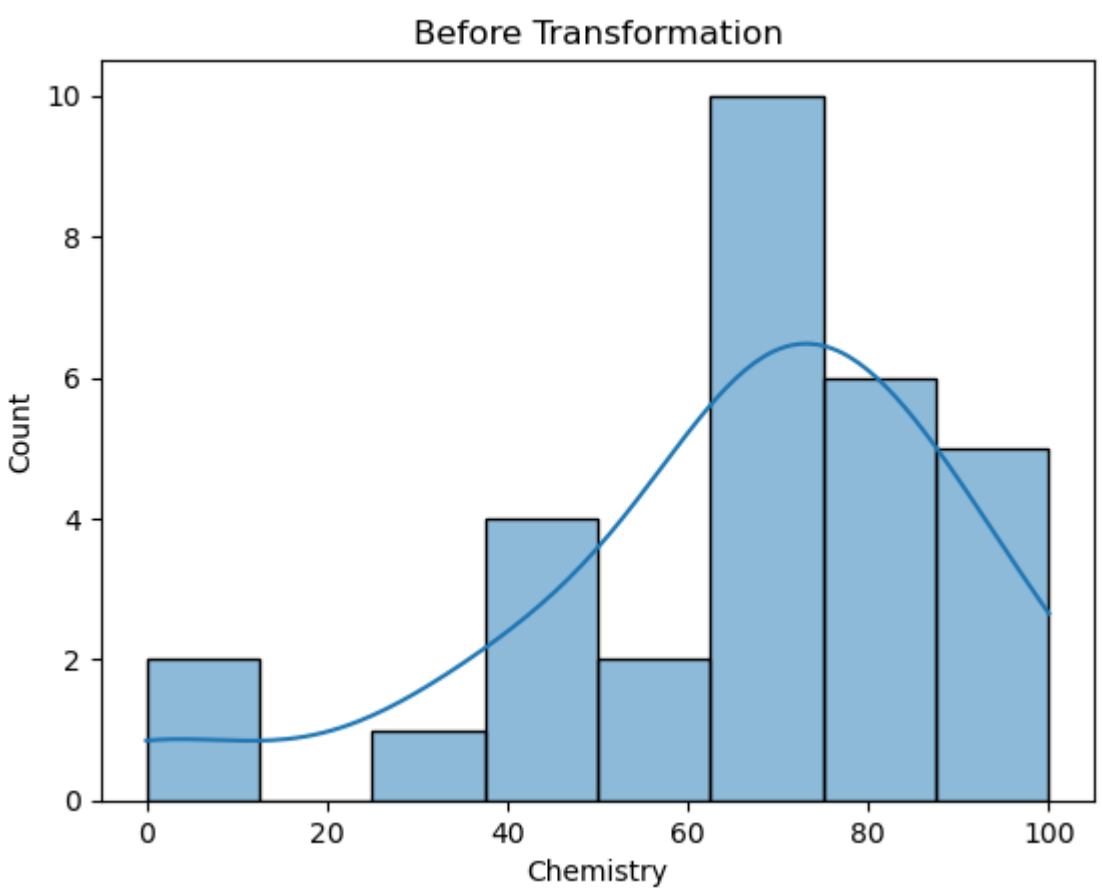
In [58]:
```
df['Chemistry'].skew()
```

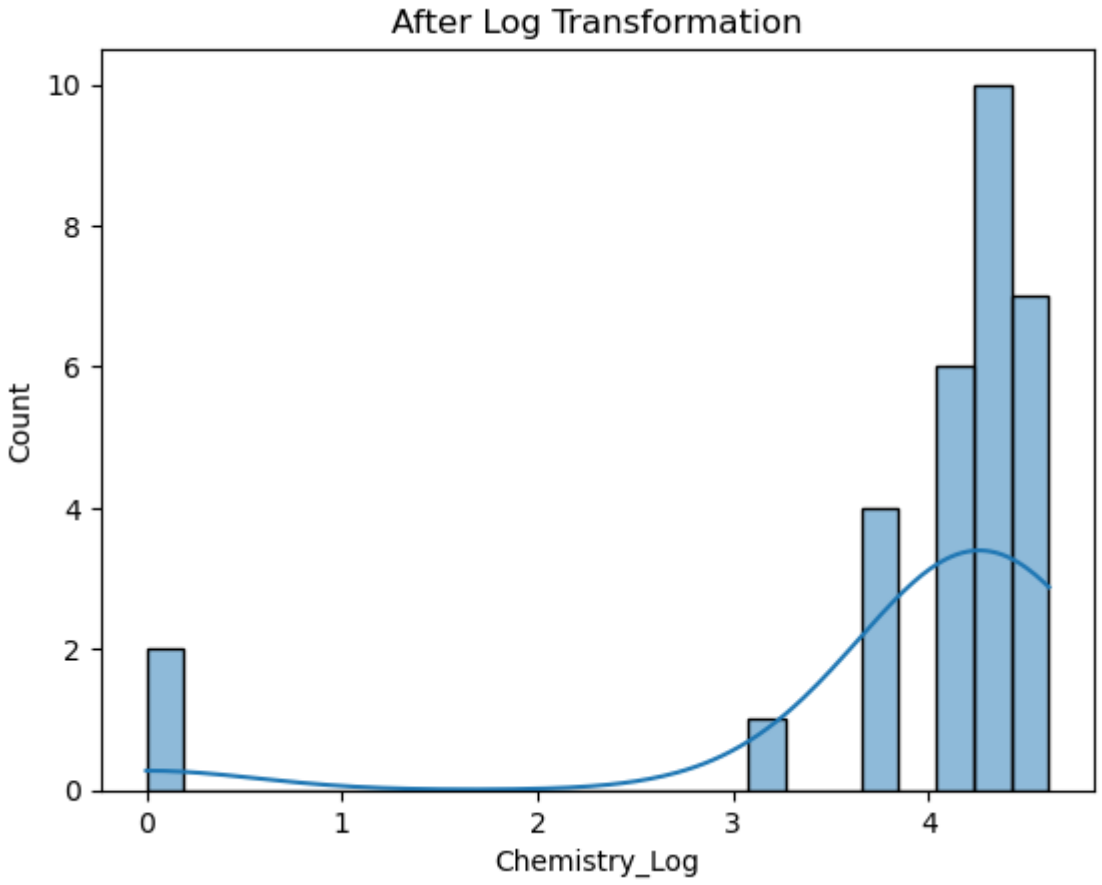
Out[58]:
```
-1.1507378639285284
```

In [60]:
```
df['Chemistry_Log'] = np.log(df['Chemistry'] + 1)
```

In [62]:
```
sns.histplot(df['Chemistry'], kde=True)
plt.title("Before Transformation")
plt.show()

sns.histplot(df['Chemistry_Log'], kde=True)
plt.title("After Log Transformation")
plt.show()
```


Before Transformation
Count
Chemistry
0
2
4
6
8
10
0
20
40
60
80
100


After Log Transformation
Count
Chemistry_Log
0
2
4
6
8
10
0
1
2
3
4

In [64]:
```
df
```

Out[64]:

	StudentId	Name	Mathematics	Physics	Chemistry	Chemistry_Log
0	1	Abhishek Podi	98.000000	78.000000	0.000000	0.000000
1	2	Abhishek Bachhan	85.000000	55.000000	45.000000	3.828641
2	3	Bikash Pandey	76.000000	96.000000	62.571429	4.152164
3	4	Anish Joshi	23.000000	100.000000	73.000000	4.304065
4	5	Abhay Sharma	72.206897	46.000000	76.000000	4.343805
5	6	Ashish Song	100.000000	66.000000	79.000000	4.382027
6	7	Ajay Cinthol	100.000000	73.206897	86.000000	4.465908
7	8	Ashutosh Song	55.000000	35.000000	65.000000	4.189655
8	9	Ayush Roll17	21.000000	81.000000	41.000000	3.737670
9	10	Abhishek Poda	88.000000	78.000000	62.571429	4.152164
10	11	Rohit Kumar	92.000000	88.000000	88.000000	4.488636
11	12	Neha Singh	81.000000	72.000000	74.000000	4.317488
12	13	Pooja Patil	77.000000	69.000000	71.000000	4.276666
13	14	Suresh Mehta	66.000000	58.000000	60.000000	4.110874
14	15	Kunal Shah	59.000000	61.000000	63.000000	4.158883
15	16	Aman Verma	100.000000	100.000000	100.000000	4.615121
16	17	Riya Desai	34.000000	42.000000	38.000000	3.663562
17	18	Vikas Rao	48.000000	39.000000	44.000000	3.806662
18	19	Sneha Nair	72.000000	75.000000	70.000000	4.262680
19	20	Mohit Jain	83.000000	84.000000	82.000000	4.418841
20	21	Kriti Malhotra	100.000000	100.000000	100.000000	4.615121
21	22	Arjun Singh	95.000000	91.000000	90.000000	4.510860
22	23	Priya Kapoor	67.000000	70.000000	68.000000	4.234107
23	24	Nitin Yadav	29.000000	28.000000	25.000000	3.258097
24	25	Shreya Bose	74.000000	73.000000	72.000000	4.290459
25	26	Deepak Gupta	13.875000	0.000000	0.000000	0.000000
26	27	Simran Kaur	90.000000	89.000000	91.000000	4.521789
27	28	Rahul Mishra	86.000000	87.000000	85.000000	4.454347
28	29	Tina Roy	60.000000	62.000000	59.000000	4.094345
29	30	Manish Tiwari	79.000000	80.000000	77.000000	4.356709

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: