In [1]:
```
import pandas as pd
import seaborn as sns
#import matplotlib.pyplot as plt
import warnings
#import numpy as np
warnings.filterwarnings('ignore')
#%matplotlib inline
```

In [2]:
```
df = pd.read_csv('C:\\Users\\Asus\\Downloads\\winequality.csv')
df.head()
```

Out[2]:

	type	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	white	7.0	0.27	0.36	20.7	0.045	45.0	170.0	1.0010	3.00	0.45	8.8	6
1	white	6.3	0.30	0.34	1.6	0.049	14.0	132.0	0.9940	3.30	0.49	9.5	6
2	white	8.1	0.28	0.40	6.9	0.050	30.0	97.0	0.9951	3.26	0.44	10.1	6
3	white	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6
4	white	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6

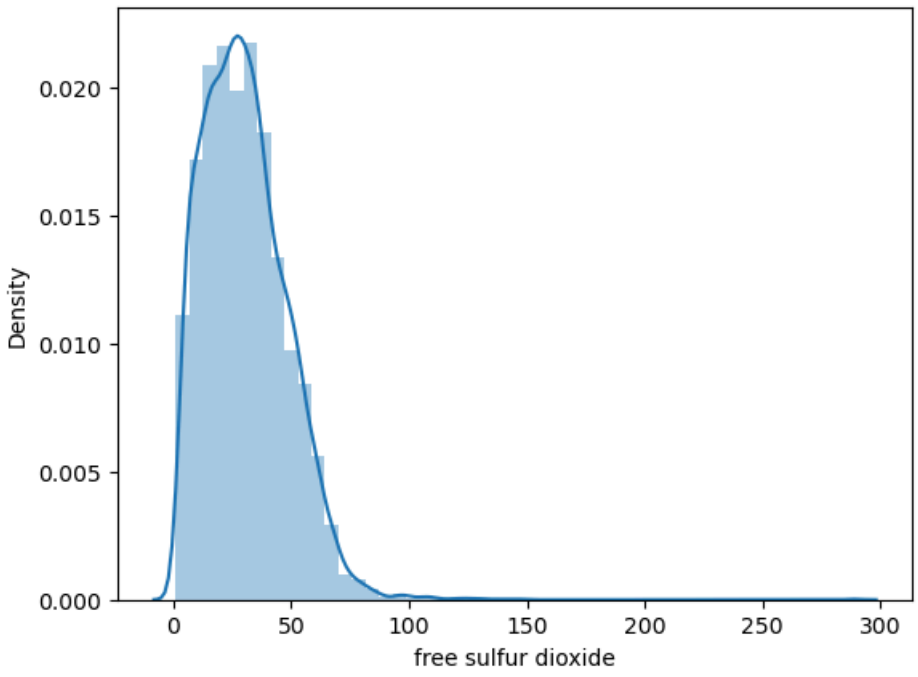
In [3]:
```
df.describe()
```

Out[3]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	s
count	6487.000000	6489.000000	6494.000000	6495.000000	6495.000000	6497.000000	6497.000000	6497.000000	6488.000000	649
mean	7.216579	0.339691	0.318722	5.444326	0.056042	30.525319	115.744574	0.994697	3.218395	
std	1.296750	0.164649	0.145265	4.758125	0.035036	17.749400	56.521855	0.002999	0.160748	
min	3.800000	0.080000	0.000000	0.600000	0.009000	1.000000	6.000000	0.987110	2.720000	
25%	6.400000	0.230000	0.250000	1.800000	0.038000	17.000000	77.000000	0.992340	3.110000	
50%	7.000000	0.290000	0.310000	3.000000	0.047000	29.000000	118.000000	0.994890	3.210000	
75%	7.700000	0.400000	0.390000	8.100000	0.065000	41.000000	156.000000	0.996990	3.320000	
max	15.900000	1.580000	1.660000	65.800000	0.611000	289.000000	440.000000	1.038980	4.010000	

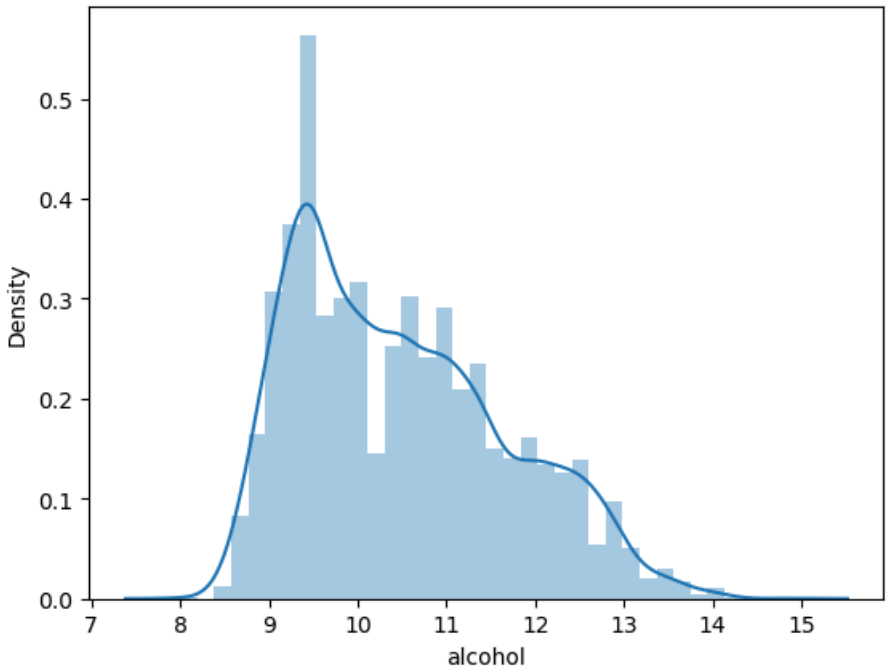
In [4]:
```
sns.distplot(df['free sulfur dioxide'])
```

Out[4]:
```
<Axes: xlabel='free sulfur dioxide', ylabel='Density'>
```



In [5]:
```
sns.distplot(df['alcohol'])
```

Out[5]: <Axes: xlabel='alcohol', ylabel='Density'>



Min-Max Scaling

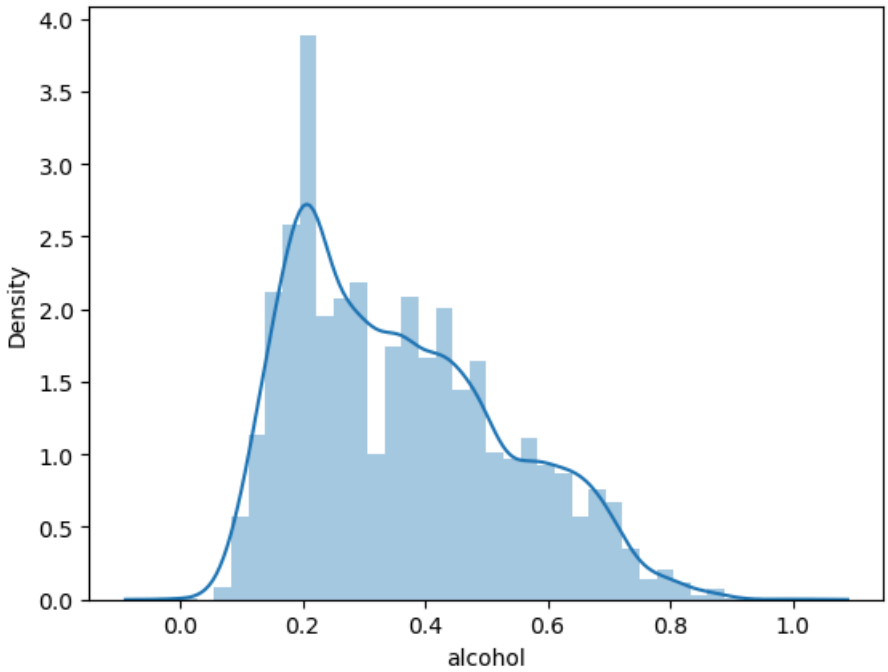
In [6]:
```
# (value - min) / (max - min)
```

In [7]:
```
df_temp = df.copy()
```

In [8]:
```
df_temp['alcohol'] = (df_temp['alcohol'] - df_temp['alcohol'].min()) / (df_temp['alcohol'].max() - df_tem
```

In [9]:
```
sns.distplot(df_temp['alcohol'])
```

Out[9]: <Axes: xlabel='alcohol', ylabel='Density'>



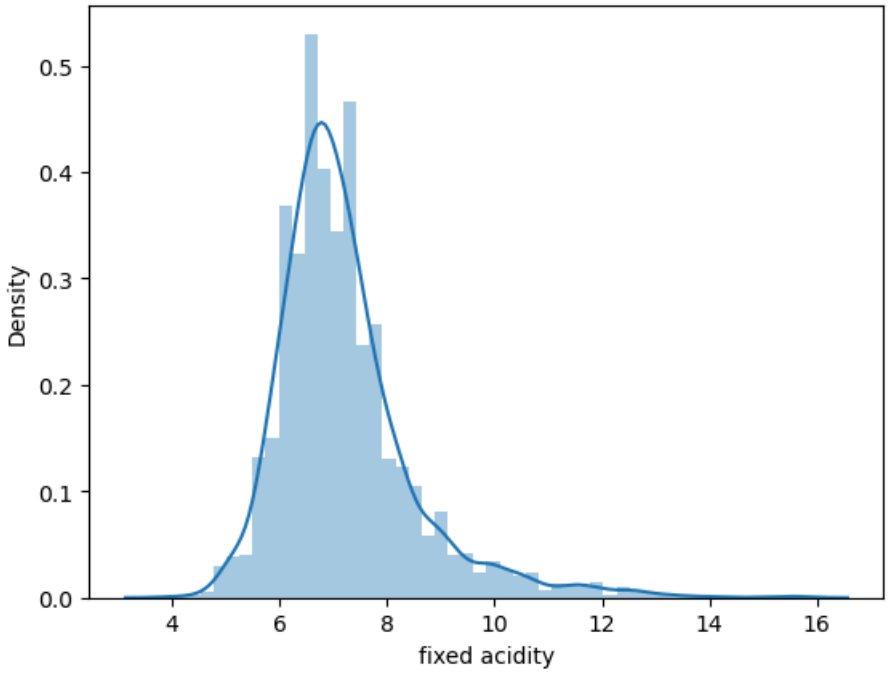
Standardization of Data (Z-Score)

In [10]:
```
## z-score method
# scaled_value = value - mean / std
```

In [11]:
```
# original_value = scaled_value * std + mean
```

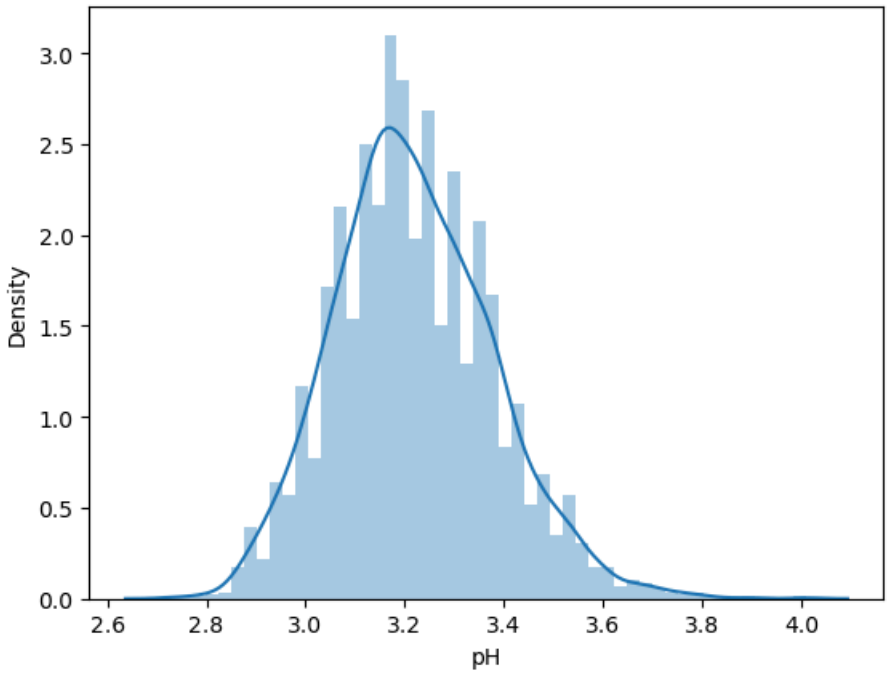
In [12]:
```
sns.distplot(df['fixed acidity'])
```

Out[12]: <Axes: xlabel='fixed acidity', ylabel='Density'>



In [13]:
```
sns.distplot(df['pH'])
```

Out[13]: <Axes: xlabel='pH', ylabel='Density'>

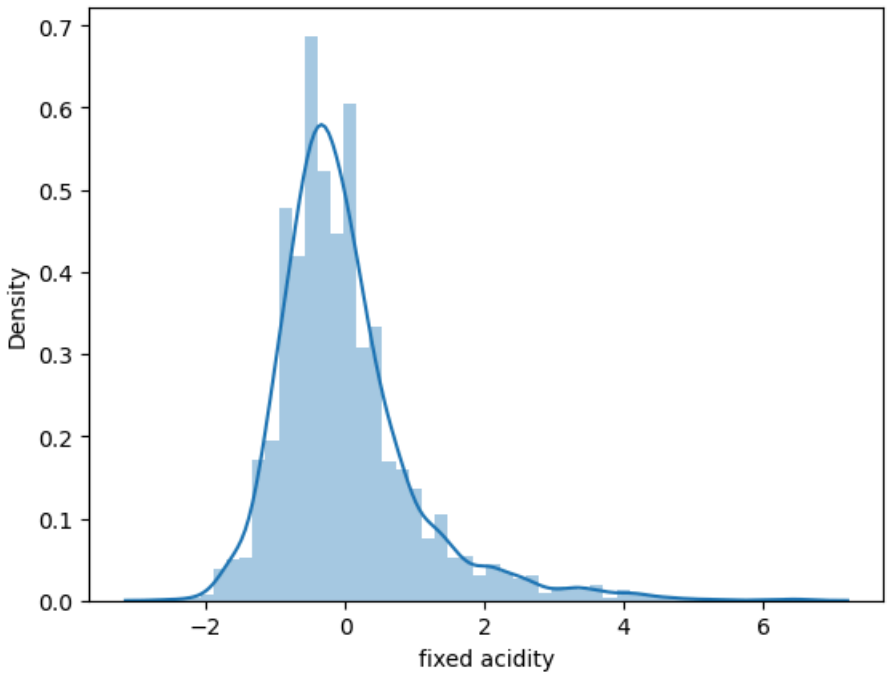


In [14]:
```
scaled_data = df.copy()
```

In [15]:
```
## apply the formula
for col in ['fixed acidity', 'pH']:
    scaled_data[col] = (scaled_data[col] - scaled_data[col].mean()) / scaled_data[col].std()
```

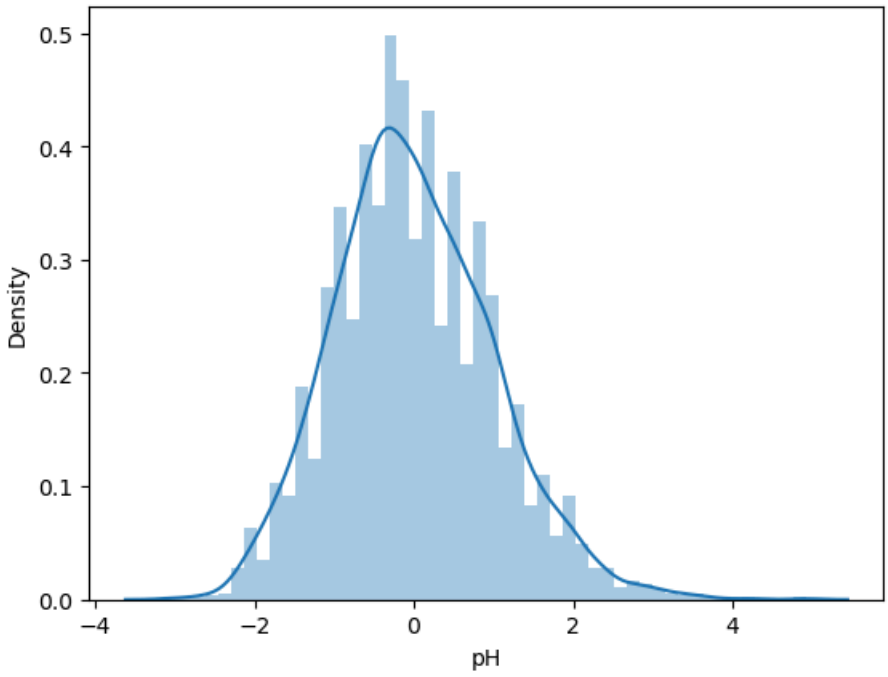
In [16]:
```
sns.distplot(scaled_data['fixed acidity'])
```

Out[16]: <Axes: xlabel='fixed acidity', ylabel='Density'>



In [17]:
```
sns.distplot(scaled_data['pH'])
```

Out[17]: <Axes: xlabel='pH', ylabel='Density'>



In [ ]:
```

```