In [33]:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

df = pd.read_csv(""C:\Users\91950\Downloads\titanic.csv"")
```

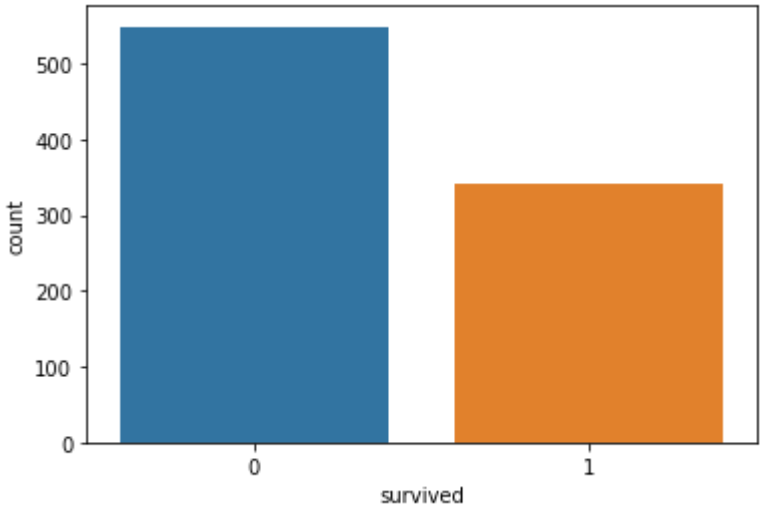
In [34]:

```
df = pd.read_csv("C:/Users/91950/Downloads/titanic.csv")
```

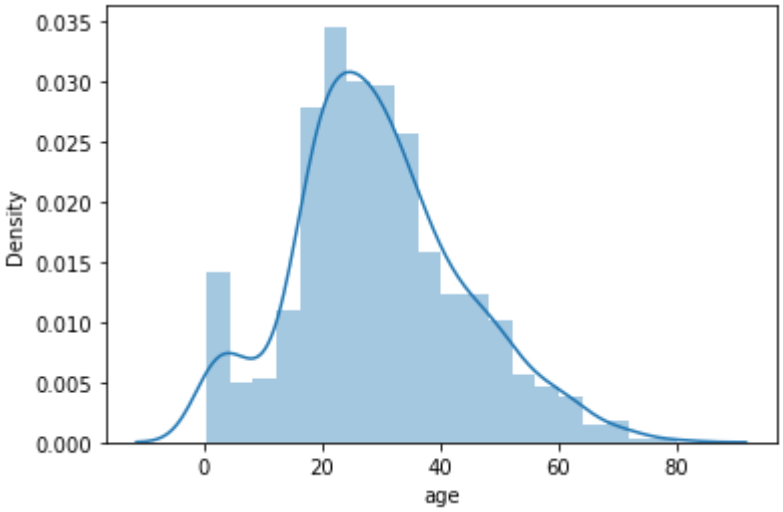
In [35]:

```
sns.countplot(df['survived'])
plt.show()

sns.distplot(df['age'])
plt.show()

sns.boxplot(df['fare'])
plt.show()
```

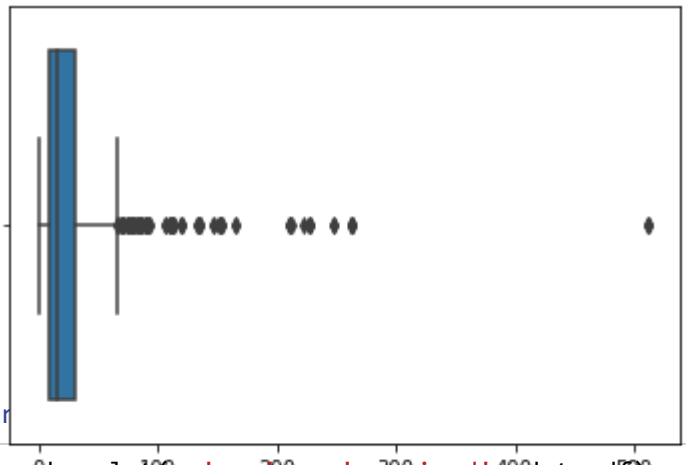
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: Futu
reWarning: Pass the following variable as a keyword arg: x. From version
0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or misinterp
retation.
  warnings.warn(
```




```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed in
a future version. Please adapt your code to use either `displot` (a figure
-level function with similar flexibility) or `histplot` (an axes-level fun
ction for histograms).
  warnings.warn(msg, FutureWarning)
```

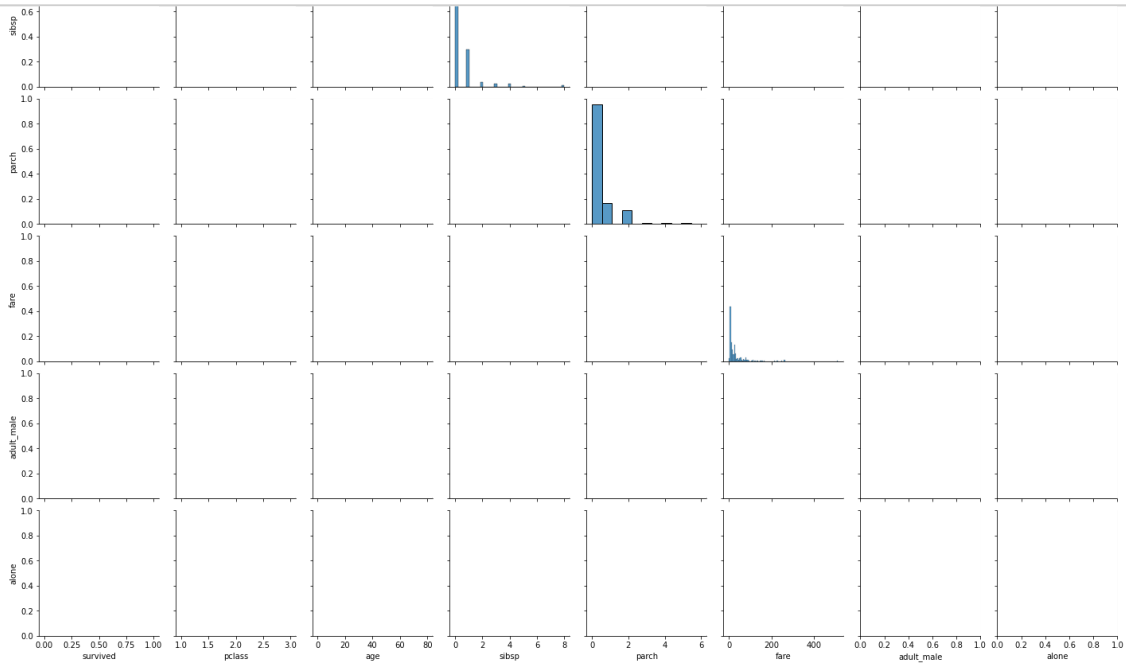


```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: Futu
reWarning: Pass the following variable as a keyword arg: x. From version
0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or misinterp
retation.
  warnings.warn(
```



In

```
sns.barplot(x='sex', y='survived', data=df)
plt.show()

sns.barplot(x='pclass', y='survived', data=df)
plt.show()

sns.lmplot(x='age', y='fare', data=df)
plt.show()

sns.pairplot(df)
plt.show()

df.describe()
```



In [ ]:

```
df['age'].fillna(df['age'].mean(), inplace=True)
df['fare'].fillna(df['fare'].mean(), inplace=True)
```

In [ ]:

```
df = df[df['age'] < 80]
df = df[df['fare'] < 1000]
```

In [ ]:

```
df['sex'] = df['sex'].astype('category')
df['embarked'] = df['embarked'].astype('category')
```

In [ ]:

```
X = df.drop('survived', axis=1)
y = df['survived']
```

In [ ]:

```
df.dropna(subset=['age', 'embarked'], inplace=True)
```

In [ ]:

```
df['fare'] = np.where((df['fare'] < lower_bound) | (df['fare'] > upper_bound),
                      df['fare'].median(), df['fare'])
```

In [ ]:

```
categorical_cols = ['sex', 'embarked']
df_encoded = pd.get_dummies(df, columns=categorical_cols, drop_first=True)
```

In [37]:

```
X = df_encoded.drop('survived', axis=1)
y = df_encoded['survived']

scaler
```

Out[37]:

```
StandardScaler()
```