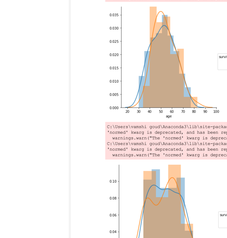
In [6]:
```
!jupyter nbconvert --to html Vamshin.ipynb
```

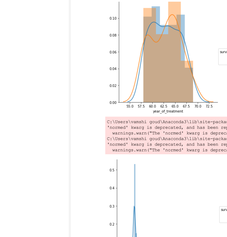
```
import pandas as pd
import numpy
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("haberman.csv")


cancer_df = pd.read_csv("haberman.csv", header=None, names=['age', 'year_of_treatment', 'positive_lymph_nodes', 'survival_status_after_5_years'])
print(cancer_df.head())
sns.FacetGrid(cancer_df, hue="survival_status_after_5_years", size=5) \
   .map(sns.distplot, "age") \
   .add_legend();
plt.show()
sns.FacetGrid(cancer_df, hue="survival_status_after_5_years", size=5) \
   .map(sns.distplot, "year_of_treatment") \
   .add_legend();
plt.show()
sns.FacetGrid(cancer_df, hue="survival_status_after_5_years", size=5) \
   .map(sns.distplot, "positive_lymph_nodes") \
   .add_legend();
plt.show()
```

```
   age  year_of_treatment  positive_lymph_nodes  survival_status_after_5_years
0   30                 64                     1                              1
1   30                 62                     3                              1
2   30                 65                     0                              1
3   31                 59                     2                              1
4   31                 65                     4                              1
```

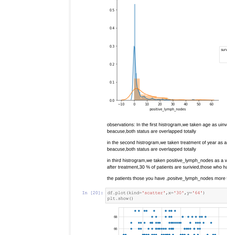
```
C:\Users\vamshi gouf\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The 'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
C:\Users\vamshi gouf\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The 'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```




```
C:\Users\vamshi gouf\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The 'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
C:\Users\vamshi gouf\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The 'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```




```
C:\Users\vamshi gouf\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The 'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
C:\Users\vamshi gouf\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The 'normed' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```



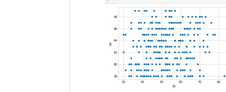
observations: in the first histogram,we taken age as univariant,in that we can not determined survival status after operation beacuse,both status are overlapped totally.

in the second histogram,we taken treatment of year as a variable,in that we can not determined survival status after operation beacuse,both status are overlapped totally

in third histogram,we taken positive_lymph_nodes as a variable,in that we can find the probability of survival status of patients after treatment,80 % of patients are survived,those who have 0-60 5% of prob operation

the patients those you have ,positive_lymph_nodes more than ,10%,their no chance for survivals more than 5 years

In [20]:
```
df.plot(kind="scatter",x="35",y="64")
plt.show()
```

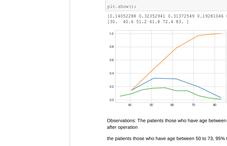


observations: form the above 1-D scatter plot we can,t determine ,survival of cancer patient,because the all points are overlaped

In [9]:
```
sns.pairplot(cancer_df, hue="survival_status_after_5_years", size=3)
plt.show()
```

In [16]:
```
sns.set_style("whitegrid")
counts, bin_edges = numpy.histogram(cancer_df['age'], bins=5,
                                    density = True)

pdf = counts/(sum(counts))
print(pdf)
print(bin_edges)
cdf = numpy.cumsum(pdf)
plt.plot(bin_edges[1:],pdf)
plt.plot(bin_edges[1:], cdf)


counts, bin_edges = numpy.histogram(cancer_df['age'], bins=10,
                                    density = True)

pdf = counts/(sum(counts))
plt.plot(bin_edges[1:],pdf)
plt.plot(bin_edges[1:], cdf)

plt.show()
```

```
[0.14051208 0.32352941 0.31372549 0.19281046 0.02941176]
[30.  40.6 51.2 61.8 72.4 83. ]
```

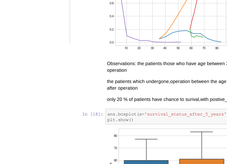


Observations: The patients those who have age between 50 to 63,only 33% of chances to survival more than 5 years after operation

the patients those who have age between 50 to 73, 99% their is no chance to survival more than 5 years after operation

In [17]:
```
import pandas as pd
import numpy
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("haberman.csv")


cancer_df = pd.read_csv("haberman.csv", header=None, names=['age', 'year_of_treatment', 'positive_lymph_nodes', 'survival_status_after_5_years'])

sns.set_style("whitegrid")
counts, bin_edges = numpy.histogram(cancer_df['age'], bins=5,
                                    density = True)

pdf = counts/(sum(counts))
print(pdf)
print(bin_edges)
cdf = numpy.cumsum(pdf)
plt.plot(bin_edges[1:],pdf)
plt.plot(bin_edges[1:], cdf)

# sirgnicie
counts, bin_edges = numpy.histogram(cancer_df['year_of_treatment'], bins=5,
                                    density = True)

pdf = counts/(sum(counts))
print(pdf)
print(bin_edges)
cdf = numpy.cumsum(pdf)
plt.plot(bin_edges[1:],pdf)
plt.plot(bin_edges[1:], cdf)


#nevisolor
counts, bin_edges = numpy.histogram(cancer_df['positive_lymph_nodes'], bins=5,
                                    density = True)

pdf = counts/(sum(counts))
print(pdf)
print(bin_edges)
cdf = numpy.cumsum(pdf)
plt.plot(bin_edges[1:],pdf)
plt.plot(bin_edges[1:], cdf)


plt.show()
```

```
[0.05229708 0.08823529 0.1565268  0.17320261 0.17973856 0.13398693
 0.13398693 0.05882353 0.02287582 0.00653595]
[30.  35.4 40.8 46.2 51.6 57.  62.4 67.8 73.2 78.6 84. ]
[0.20588235 0.29915033 0.18464052 0.1751634  0.09803922 0.10130719
 0.03267974 0.01633987 0.00326797 0.00653595]
[58.  59.2 60.4 61.6 62.8 64.  65.2 66.4 67.6 68.8 70. ]
[0.77124183 0.59803922 0.05882353 0.02614379 0.02941176 0.00653595
 0.00326797 0.         0.00326797 0.00326797]
[ 0.   5.2 10.4 15.6 20.8 26.  31.2 36.4 41.6 46.8 52. ]
```

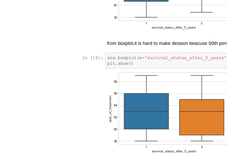


Observations: the patients those who have age between 30 to 50,they have 78% chance to live,more than 5 years after operation

the patients which undergone,operation between the age of 50 to 60 their is 18 to 19% of chance to survival more than 5 years after operation

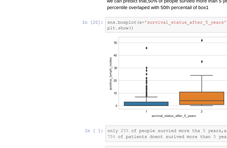
only 20 % of patients have chance to survival,with positive_lymph_nodes,whose age lies between 30 to 60

In [18]:
```
sns.boxplot(x="survival_status_after_5_years",y="age", data=cancer_df)
plt.show()
```



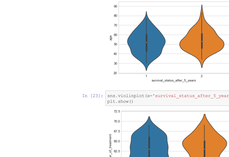
from boxplot,it is hard to make derision because 50th percentile of both box plots are overlaped

In [19]:
```
sns.boxplot(x="survival_status_after_5_years",y="year_of_treatment", data=cancer_df)
plt.show()
```



we can predict that,50% of people survied more than 5 year,those you done operation between 60 to 66 year ,beacuse 75th percentile overlaped with 50th percentail of box1

In [20]:
```
sns.boxplot(x="survival_status_after_5_years",y="positive_lymph_nodes", data=cancer_df)
plt.show()
```

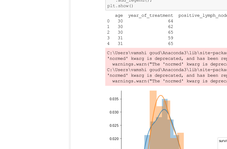


In [ ]:
```
only 25% of people survied more tha 5 years,after operation with positive_lymph_nodes
75% of patients doest survived more than 5 year after operation with positive_lymph_nodes
```

In [22]:
```
sns.violinplot(x="survival_status_after_5_years",y="age", data=cancer_df,size=10)
plt.show()
```

In [23]:
```
sns.violinplot(x="survival_status_after_5_years",y="year_of_treatment", data=cancer_df,size=10)
plt.show()
```


In [24]:
```
sns.violinplot(x="survival_status_after_5_years",y="positive_lymph_nodes", data=cancer_df,size=10)
plt.show()
```



In [ ]:
```

```