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

os.getcwd()
```

Out[1]: 'C:\\Users\\Simha\\Notebooks'

In [3]:
```
'''
Attribute Information:

Age of patient at time of operation (numerical)
Patient's year of operation (year - 1900, numerical)
Number of positive axillary nodes detected (numerical)
Survival status (class attribute) 1 = the patient survived 5 years or longer 2 = the patient died within 5 year



objective ::

Find out factors that decide the Probability of a patient to survive successfully after operation



'''



haberman=pd.read_csv('haberman.csv')


print(haberman.shape)

haberman.head()
```

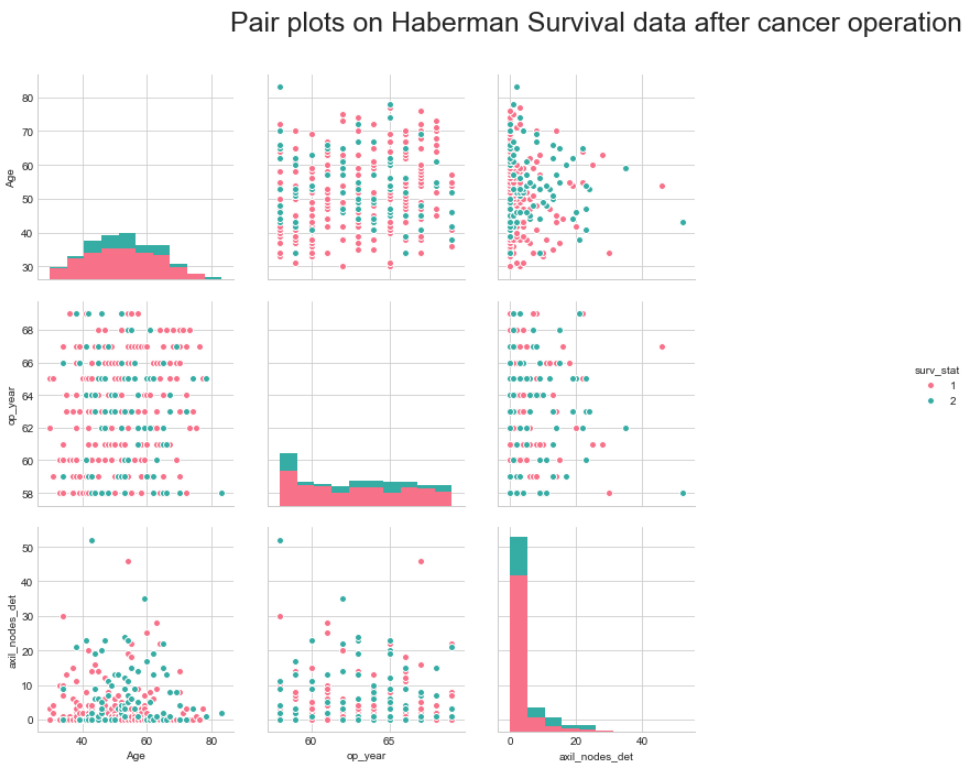
```
(305, 4)
```

Out[3]:

	Age	op_year	axil_nodes_det	surv_stat
0	30	62	3	1
1	30	65	0	1
2	31	59	2	1
3	31	65	4	1
4	33	58	10	1

Pair plots

In [47]:
```
sns.set_style('whitegrid');
sns.pairplot(haberman,vars=['Age','op_year','axil_nodes_det'],hue='surv_stat',size=3,palette='husl');
plt.title('Pair plots on Haberman Survival data after cancer operation\n\n\n\n\n\n\n\n\n\n',fontsize=25)
plt.show();
#fig=plt.figure()
#fig.suptitle('Pair plots on Haberman Survival data after cancer operation',fontsize=20)
```



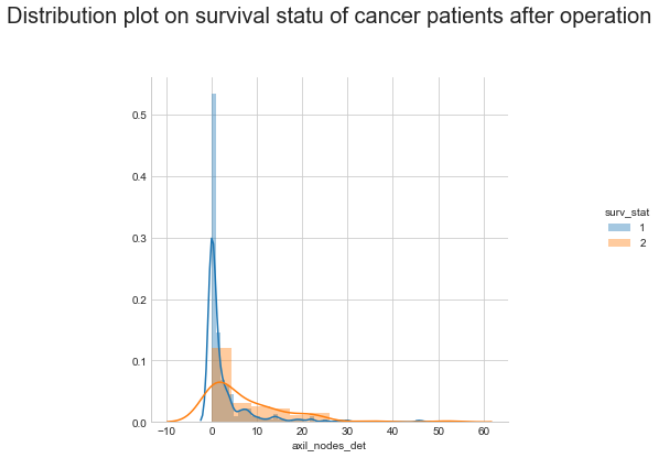
From above pair plot we can tell that

1. People with age between 40 and 60 years are having more probability of getting detected with auxilary nodes of cancer .
2. The very next bracket of age is 60 to 80 years where getting detected with auxilary nodes of cancer is high.

Conclusion :

Roughly 75% of total breast cancer patients are having age as greater than 40 years.

In [28]:
```
# Probability Distribution Function
print('\n')
sns.FacetGrid(haberman,hue='surv_stat',size=5).map(sns.distplot,'axil_nodes_det').add_legend();
plt.title('Distribution plot on survival statu of cancer patients after operation\n\n',fontsize=20)
plt.show();
```



Conclusion : Shows chances to live more than 5 years are high after operation if, number of auxilary nodes that spotted are lessthan or equal to 3

In [6]:
```
#Mean
print(np.mean(haberman['Age'])) #prints mean value of column Age in dataset
print(np.mean(haberman['op_year']))
print(np.mean(haberman['axil_nodes_det']))
print(np.mean(haberman['surv_stat']))
```

```
52.5311475409836
62.84918032786885
4.036065573770492
1.2655737704918033
```

In [19]:
```
print(np.std(haberman['Age'])) #standard deviation of colmn Age in Dataset
```

```
10.726396748570311
```

In [10]:
```
print(np.percentile(haberman['Age'],np.arange(0,100,25))) #Percentiles in Quantiles for column Age
```

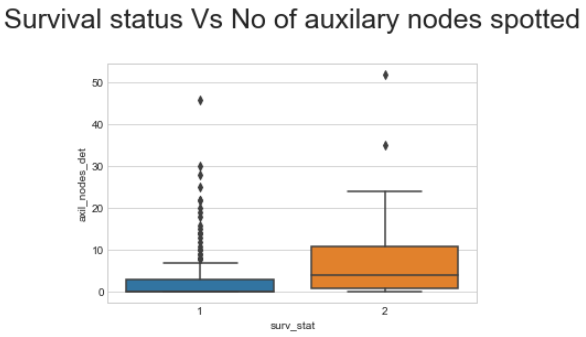
```
[30. 44. 52. 61.]
```

In [13]:
```
from statsmodels import robust

print(robust.mad(haberman['Age'])) #Median Absolute Deviation for column Age in dataset
```

```
11.860817748044816
```

Box plot

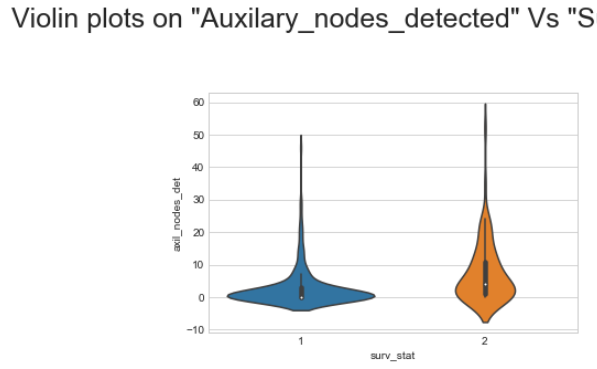
In [33]:
```
sns.boxplot(x='surv_stat',y='axil_nodes_det',data=haberman); #Boxplots
plt.title("\nSurvival status Vs No of auxilary nodes spotted\n",fontsize=25)

plt.show();
```



Conclusion : Shows that 75% people survived more than 5 years after operation if they spotted with <=4 auxilary nodes of cancer in breast. Also people who got detected with more than 5 auxilary nodes had died within 5 years after operation.

Violin plot

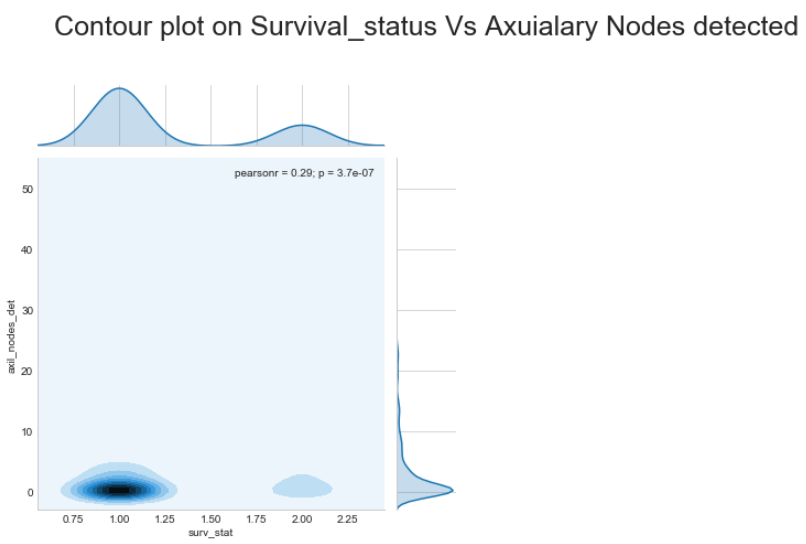
In [39]:
```
sns.violinplot(x='surv_stat',y='axil_nodes_det',data=haberman);
plt.title('\n Violin plots on "Auxilary_nodes_detected" Vs "Survival_status" \n \n ',fontsize=25)

plt.show();
```



Conclusion : From the above violin polt, blue colour plot say that people who survived more than 5 years after operation are having auxilary nodes detected as <=5 before operation. Where asorange plot in figure tells us people died within 5 years if they got detected with more than 5 auxilary nodes on breast

Multivariate probability density, Contour plot

In [45]:
```
sns.jointplot(x='surv_stat',y='axil_nodes_det',data=haberman,kind='kde');
plt.title('\n Contour plot on Survival_status Vs Axuialary Nodes detected \n\n\n\n',fontsize=25)
plt.show();
```



Conclusion : Shows that 80% of people survived more than 5 years after operation , only if they spotted with <=4 auxilary nodes of cancer in breast. Also people who got detected with more than 5 auxilary nodes had died within 5 years after operation.

Conclusion

After analysing all charts we can conclude from this dataset is , The survival status of 80% of people is more than 5 years afer the operation if and only if the number of auxiliary nodes got detected in their Breast is les than or equal to 5.