

Aim - To implement statistical concept on Haberman survival dataset.

Question 1

Checking Dataset for null values.

```
In [13]: df.isnull()
```

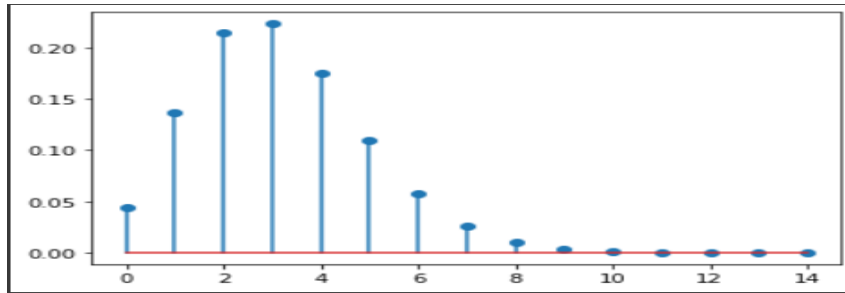
```
Out[13]:
```

	Age	op_year	axil_nodes	surv_status
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...
301	False	False	False	False
302	False	False	False	False
303	False	False	False	False
304	False	False	False	False
305	False	False	False	False

306 rows × 4 columns

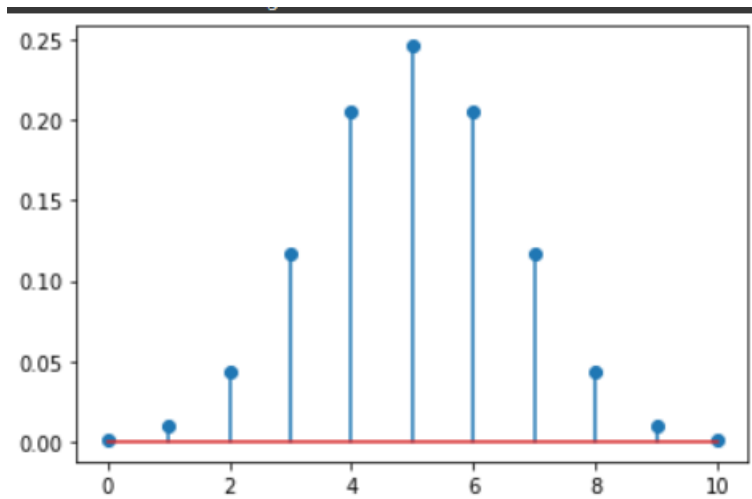
Question 2. Imagine you have a machine learning model deployed in the cloud and receiving requests from your customers in real-time. How much cloud resources do you need to pay for in order to be 99% sure you can serve all the traffic that arrives at the model in any one-minute period? (Note: 3.3 requests on average based on your traffic data). Draw the distribution using python. Hints: Poisson distribution

```
#Q2
from scipy.stats import poisson
rate = 3.13
probs = [poisson.pmf(i,rate) for i in range(15)]
plt.stem(list(range(15)),probs)
```



Question 3. What is the probability of observing different numbers of heads in 10 tosses with a fair coin? Find which distribution will get apply and plot it. Write a python script to draw the distribution. Hints: Binomial Distribution.

```
#Q3
from scipy.stats import binom
num = 10
head = 0.5
prob = [binom.pmf(i,num,head) for i in range(11)]
plt.stem(list(range(11)),prob)
```



Conclusion - Statistical concepts implement successfully.