GROUP ASSIGNMENT # 04

STATISTICS AND PROBABILITY THEORY

NAMES	REGISTRATION
SUDAIS KHAN	FA20-BSE-042
ASADULLAH	FA20-BSE-028
UBAID UR REHMAN	FA20-BSE-012
SECTION	BSE-5B

- **Submitted To Sir Atta Ullah**

Question 1: Discuss the following probability distributions with applications.

1: Binomial Distribution:

Binomial distribution is a probability distribution used in statistics that summarizes the likelihood that a value will take one of two independent values under a given set of parameters or assumptions.

The underlying assumptions of binomial distribution are that there is only one outcome for each trial, that each trial has the same probability of success, and that each trial is mutually exclusive, or independent of one another.

Applications:

- Number of Side Effects from Medications
- Number of Fraudulent Transactions
- Number of Spam Emails per Day

2: Poison Distribution:

A Poisson distribution is a discrete probability distribution. It gives the probability of an event happening a certain number of times (k) within a given interval of time or space.

The Poisson distribution has only one parameter, λ (lambda), which is the mean number of events. The graph below shows examples of Poisson distributions with different values of λ.

Applications:

- Horse kick deaths
- Text messages per hour
- Website visitors per month
- Influenza cases per year

3: Exponential Distribution:

The exponential distribution is a continuous probability distribution that often concerns the amount of time until some specific event happens. It is a process in which events happen continuously and independently at a constant average rate.

The exponential distribution has the key property of being memoryless. The exponential random variable can be either more small values or fewer larger variables.

Applications:

- Predict the time when an Earthquake might occur.
- Call Duration.
- Life Span of Electronic Gadgets.
- Cars Passing per Minute.

4: Normal Distribution:

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

Applications:

- Student's Average Report
- Technical Stock Market
- Tossing a Coin
- Rolling a Dice

Question 2: (Using Python)

a) Fit binomial distribution on Gender data and what percent of chance that next BCS class in SP23 will consist on (1) 20 females (2) at least 15 males.

Description: In this step Reading the from csv file.

```
[1] from google.colab import drive
    drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
[2] import pandas as pd
    import numpy as numpy
    dataset = pd.read_csv('/content/gdrive/My Drive/Assign4.csv')
```

```
[3] dataset.head()
```

	S,NO	GENDER	QUIZ\n(X1)	ASSIGN\n(X2)	MID\n(X3)	FINAL\n(Y)	TOTAL	LG	GP
0	1	F	11.3	9.3	23	32.9	77	B	3
1	2	M	12.3	8.3	19.5	28.6	69	C+	2.3
2	3	F	9.2	7.7	8	25.7	51	D	1.3
3	4	M	12.9	8.7	21	25.7	68	C+	2.3
4	5	F	11.3	8.7	21.5	4.0	82	B+	3.3

Description: In this step Fit binomial distribution on Gender for Females.

```
[8] # For females
    from scipy.stats import binom

    #calculate binomial probability
    binom.pmf(k=20, n=69, p=0.2)
```

```
0.02163150750277834
```

Description: In this step Fit binomial distribution on Gender for Males.

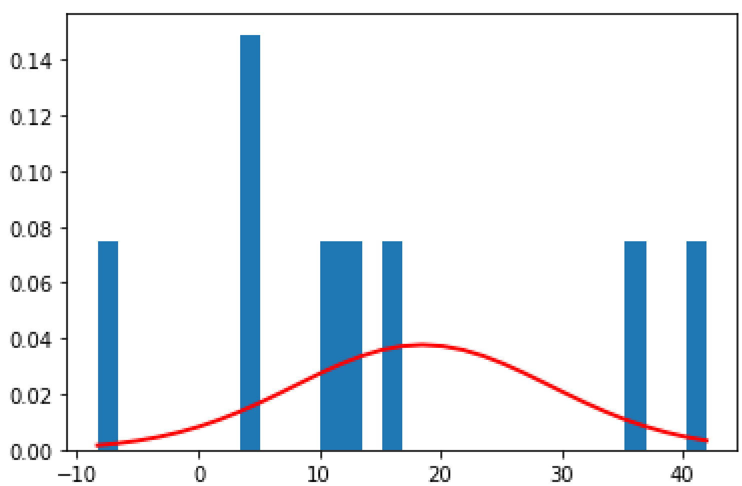
```
# For Males
from scipy.stats import binom
#calculate binomial probability
binom.pmf(k=15, n=69, p=0.2)
```

```
0.10859219093025141
```

b) Fit normal distribution on final marks Y and what percent of chance that you will get marks predicted marks calculated in assignment number 2.

Description: In this Step Fit Normal Distribution on FINAL MARKS.

```
[19] from scipy.ndimage.measurements import mean
     from pandas.io.parsers.readers import read_csv
     import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import statistics
     data = read_csv("/content/gdrive/My Drive/Assign4.csv")
     final=data['FINAL\n(Y)'].tolist()
     lst1=statistics.mean(final)
     lst2=statistics.stdev(final)
     mu,sigma=lst1,lst2
     s=np.random.normal(mu,sigma,8)
     count,bins,ignored=plt.hist(s,30,density=True)
     plt.plot(bins,1/(sigma*np.sqrt(2*np.pi))*np.exp(-(bins-mu)**2/(2*sigma**2)),linewidth=2,color='r')
     plt.show()
```



c) Fit exponential distribution on Z, considering it required time (hour) to solve the final exam paper. and what percent of chance that you will solve final paper within 3 hour.

Description: In this step fit exponential distribution on GP Column.

[24]
```
from numpy.random import exponential
GP=data['GP'].tolist()
lst3=len(GP)
random_number=exponential(scale=3,size=lst3)
print(random_number)
```

```
[2.30340342e+00 4.91635500e+00 3.57710439e+00 9.66468917e-01
 2.15902411e-01 2.87522714e+00 2.25588554e+00 4.29232749e-01
 1.04053886e+01 1.50973735e+00 6.66107878e+00 4.10450916e+00
 1.13292645e-03 1.03074650e+01 2.84708148e+00 1.22118143e+00
 2.24937801e+00 5.62137496e+00 1.62681776e-01 5.78407960e+00
 4.03357956e+00 2.47931935e-01 2.98796944e+00 1.26038453e-01
 2.67324948e-01 3.09828741e+00 1.31143273e+00 3.68570914e+00
 4.52018948e+00 1.79337898e+00 3.13918248e+00 5.49745545e-01
 2.85689467e+00 6.99667574e-01 2.42407432e+00 2.48309252e+00
 4.16117790e+00 1.03595406e+01 1.89471516e+00 1.01061820e+00
 2.04085019e+00 2.48450302e-01 1.15945233e+00 5.84045118e+00
 1.04327792e+00 4.66785418e+00 3.41933573e+00 2.83987791e-01
 9.30855315e-01 2.63805039e+00 4.14773026e+00 3.77148933e+00
 2.61028053e-01 4.07188517e+00 2.29505104e-01 1.16004178e+00
 4.49904340e+00 4.25500109e+00 3.27759711e+00 2.11753394e+00
 1.68855412e+00 3.00148389e+00 6.59050759e-01 3.92818470e+00
 2.97112922e-02 1.99826444e+00 5.90131468e-01 1.00384443e+00
 1.54712369e+00]
```