HOMEWORK 2

Kalantarpour, Cyrus

[COMPANY NAME] [Company address]

Part 1:

Since this is a Markov Process we have:

$$E[S] = \frac{60}{20} * 2.5 = 7.5\ min$$

$$E[S^2] = \frac{60}{20} * 2.5^2 = 18.75\ min^2$$

$$\rho = \lambda E[S] = 22.5$$

To run the simulation I used Python Numpy `np.random.poisson(20, 1000)`. For the average time of geeting answer, I am not sure whether it is constant (2.5 min) or not. Therefore, I used a normal distribution with mean 2.5 and var 0.1.

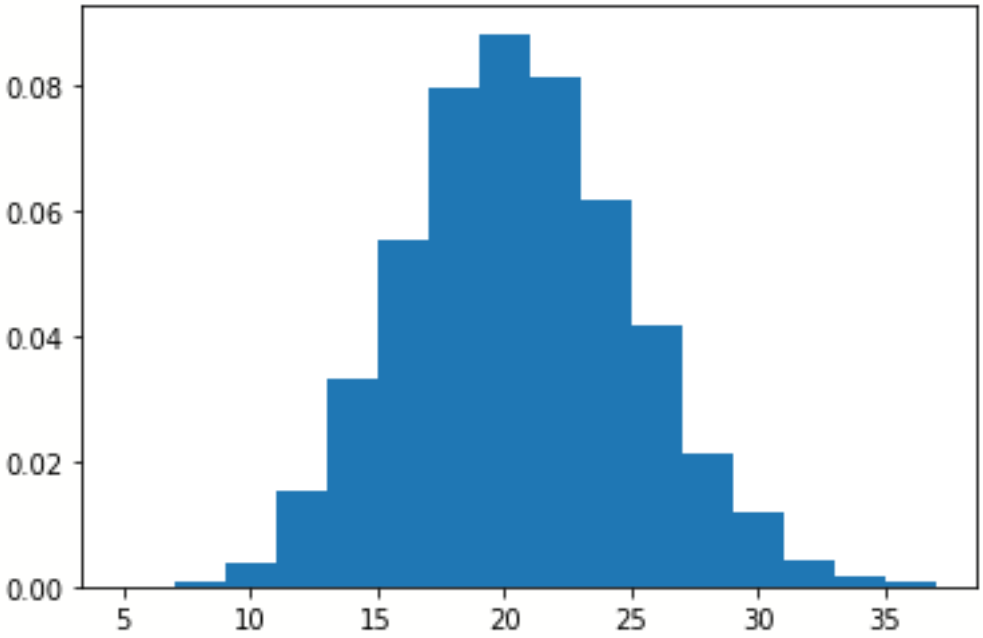

Figure 1: Histogram Plot of the 1000 points from Poisson distribution. $\lambda = 20$

Part 2:

I used Pandas library in Python to load 'Crime.xlsx'. Since it contains 3 sheets, I remove the sheets that contains no data. Then I used pandas.read_excel method to read the 'Crime.xlsx' (Figure 2). Next Step I used `sklearn.preprocessing` standard scaler to scale the data. Afterwards I applied the PCA to data. To do so, I divide the crime rate column' value into three Low(from 34.2 to 55) Medium(from 55 to 89.2) and high(from 89.2 to 144) crime rates to be able to plot the PCA result using Matplot lib.

I plotted a 2d PCA first and second component representation. (Figure 4)

```
1 import pandas as pd
2 Crime = pd.read_excel("Crime.xlsx")
3 Crime
```

	Unnamed: 0	Age	S	Ed	Ex0	Ex1	LF	M	N	NW	U1	U2	W	X	Crime Rate
0	NaN	151	1	91	58	56	510	950	33	301	108	41	394	261	79.1
1	NaN	143	0	113	103	95	583	1012	13	102	96	36	557	194	163.5
2	NaN	142	1	89	45	44	533	969	18	219	94	33	318	250	57.8
3	NaN	136	0	121	149	141	577	994	157	80	102	39	673	167	196.9

Figure 2: Reading Crime data using Pandas.

	principal component 1	principal component 2	Crime Rate
0	3.958211	1.491759	79.1
1	-1.023794	-0.699377	163.5
2	3.938648	-0.020112	57.8
3	-3.615432	2.439577	196.9
4	-1.849880	-1.338246	123.4

Figure 3: First two principal components table

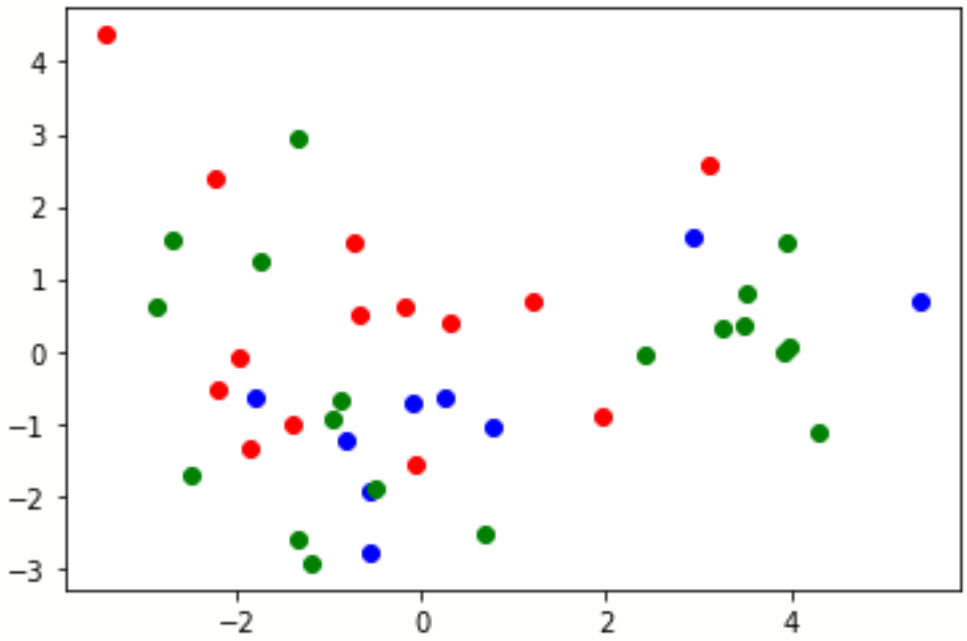

Figure 4: plot of the figure 3' fists and second columns. Each point represents a row of the principal components. Coloring is done according to the associated Crime-Rate value of row of the principal components. The Blue, Green and red dots are the representation of Low, medium, and high crime rates principal components. The X axis is the first component and the Y axis is the Second component.

Part 3:

In order to read the data, I used Pandas (Figure 5). Then I used `train_test_split` from `sklearn` and separated data into train and test set with test /train ratio of 0.2 (Figure 6). Afterward I used mean square error metrics to compute the model error which is `0.21281266507699947`.

	Recu	test	moyenne	reputation
0	0	380	3.61	3
1	1	660	3.67	3
2	1	800	4.00	1
3	1	640	3.19	4
4	0	520	2.93	4

Figure 5: Reading the data via Pandas

```
y_prediction =  LR.predict(x_test)
y_prediction
```

```
array([ 0.38655448,  0.4887523 ,  0.42482326,  0.17143575,  0.28750036,
        0.25880539,  0.02298134,  0.40186272,  0.09615412,  0.51933141,
        0.49814421,  0.49134121,  0.20685718,  0.096486  ,  0.36917657,
        0.33216552,  0.0930496 ,  0.42201327,  0.23536479,  0.37106253,
        0.42482326,  0.22456698,  0.47832555,  0.22475251,  0.2219407 ,
        0.1807903 ,  0.3712836 ,  0.31219869,  0.19398968,  0.49049009,
        0.48631155,  0.0990731 ,  0.28069555,  0.10550867,  0.30136351,
        0.3328293 ,  0.46372208,  0.20023608,  0.41055168,  0.18984851,
        0.3719492 ,  0.40552202,  0.23639963,  0.60452238,  0.22353396,
```

Figure 6: The Multi variable Regression Prediction pre-view