Assignment - 2

TITLE:
Normal Distribution

PROBLEM STATEMENT:
Plot the Normal Distribution for class test results of a particular subject. Identify the skewness and kurtosis.

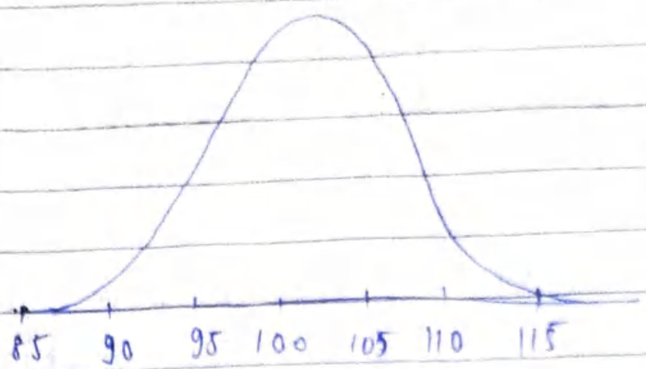
OBJECTIVES:
To get detailed approach of simulation, estimation and visualization of statistical data.

OUTCOMES:
Apply appropriate statistical concepts and skills to solve problems in both familiar and unfamiliar situations.

PREREQUISITES:
→ Concept of data distribution

THEORY:

1) Normal Distribution
The normal distribution is an important continuous distribution because a good no. of random variables occuring in practice can be approx. to it. If a random variable is affected by many independent causes, and effect of each cause is not overwhelmingly large compared to other effects, then it will closely follow normal distribution.



A normal distribution with Mean 100 and standard deviation 5.

For a normal distribution with mean μ and standard deviation s, the $f(x)$ is given by complicated formula:

$$f(x \mid \mu \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

→ Properties of Normal Distribution

If several independent random variables are normally distributed, then their sum will also be normally distributed. The mean of sum will be sum of all the individual means, and by virtue of independence, the variance of sum will be sum of individual variances.

We can write this is algebraic form as

$$E(S) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

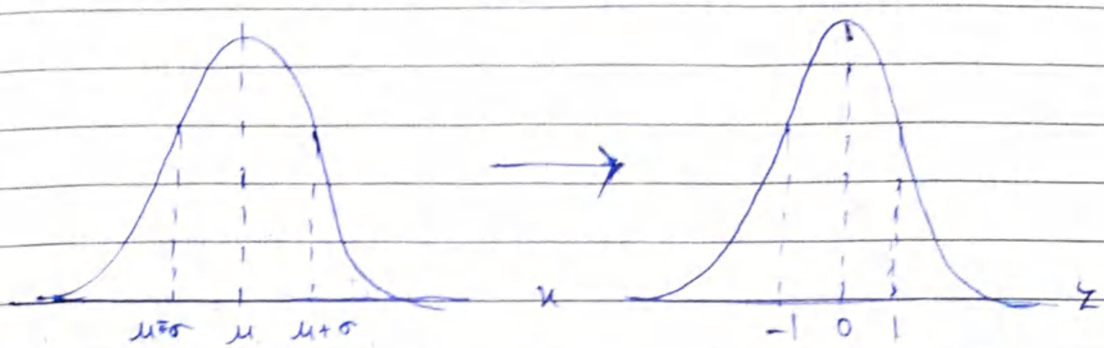
and

$$V(S) = V(X_1) + V(X_2) + \cdots + V(X_n)$$

→ The standard Normal Distribution

Standard Normal distribution is special case of Normal distribution when $\mu = 0$ and $\sigma = 1$. For any normal distribution, we can convert it into Standard Normal distribution using the formula.

$$Z = \frac{x - \mu}{\sigma}$$



2) Skewness

In addition to measures to location, such as the mean or median and measures of variation such as variance or standard deviation, two more attributes of frequency distribution of a data set many be of interest to us.
These are skewness and kurtosis.
Skewness is a measure of degree of asymmetry of a frequency distributed.

Date: / /20

3) Kurtosis

Kurtosis is a measure of peakedness of a distribution. The larger the kurtosis, the more peaked will be distribution. The kurtosis is calculated and reported either as an absolute or relative value. Absolute kurtosis is a positive no.

$$\text{Relative kurtosis} = \text{Absolute Kurtosis} - 3$$

4) Plot the Normal Distribution

→ Histogram

A histogram visualises distribution of data over a continuous interval. Each bar in a histogram represents the tabuted frequency at each interval / bin.

```
# setting the ranges and no. of intervals
range = (0, 100)
bins = 10

# plotting a histogram
plt.hist(df['History-Marks'] bins, range,
     color='blue', histype='bar', rwidth=0.8)

# x-axis label
plt.xlabel('Marks')
# frequency label
plt.ylable('Frequency')
# plot title
plt.title('histogram: Marks in history')
plt.show()
```

→ KDE plots

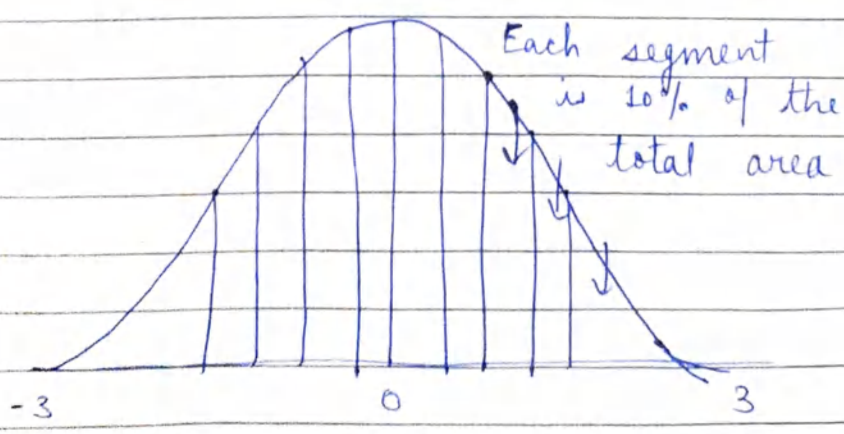
Histogram results can vary widely if you set different no. of bins and simply change start and end values of a bin. To overcome this, we can make use of density function. A density plot is a smoothed, continuous version of histogram estimated from the data. The most common form of estimation is known as kernel density estimation. (KDE).

→ Q – Q Plot

Quantiles are cut points dividing the range of a probability distribution of continuous intervals with equal probabilities or dividing observations in a sample in the same way.

→ 2 quantile is Median.
→ 4 quantile is Quartile.
→ 10 quantile is Decile.
→ 100 quantile is Percentile.

→



Pg.No :
Date : / /20

10 quantile will divide the Normal Distribution into 10 parts each having 10% of the data points.

CONCLUSION:

Hence plotted the Normal Distribution for class test result of a particular subject.

AIML Assignment2.ipynb - Colab

colab.research.google.com/drive/1WFRouddXKbWamW5jQWnt-NjQPXJM2ju

Apps HackerRank W3Schools Elearn Codeforces Django ERDPlus LeetCode Tinkercad Coursera My Captain Google Tech Dev G... Reading list

AIML Assignment2.ipynb

File Edit View Insert Runtime Tools Help Last saved at 1:30 PM

Comment Share

Files

Connecting to a runtime to enable file browsing.

+ Code + Text Allocating Editing

Plot the Normal Distribution for class test result of a particular subject. Identify the Skewness and Kurtosis

```
[ ] from scipy.stats import skew, kurtosis
    import matplotlib.pyplot as plt
    import scipy.stats as stats
    import pandas as pd
    import numpy as np
```

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

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
[ ] df = pd.read_csv("/content/drive/MyDrive/data/StudentsPerformance2.csv")
    x = df['math score']
    x =x.to_numpy()[:15]
    x.sort()
```

```
[ ] mean= np.mean(x)
    sd = np.std(x,ddof =1)
    median = np.median(x)
    fit = stats.norm.pdf(x,mean,sd)
```

```
[ ] plt.hist(x, density = True, color ="yellow",ec ="white")
    plt.plot(x, fit,"go:")
    plt.title("Math Scores")
```

Waiting for colab.research.google.com...

Type here to search 24°C Light rain ENG 13:55 13-01-2022

AIML Assignment2.ipynb - Colab

colab.research.google.com/drive/1WFRouddXKbWamW5jQWnt-NjQPXJM2ju

Apps HackerRank W3Schools Elearn Codeforces Django ERDPlus LeetCode Tinkercad Coursera My Captain Google Tech Dev G... Reading list

AIML Assignment2.ipynb

File Edit View Insert Runtime Tools Help Last [redacted] at [redacted] PM

Comment Share

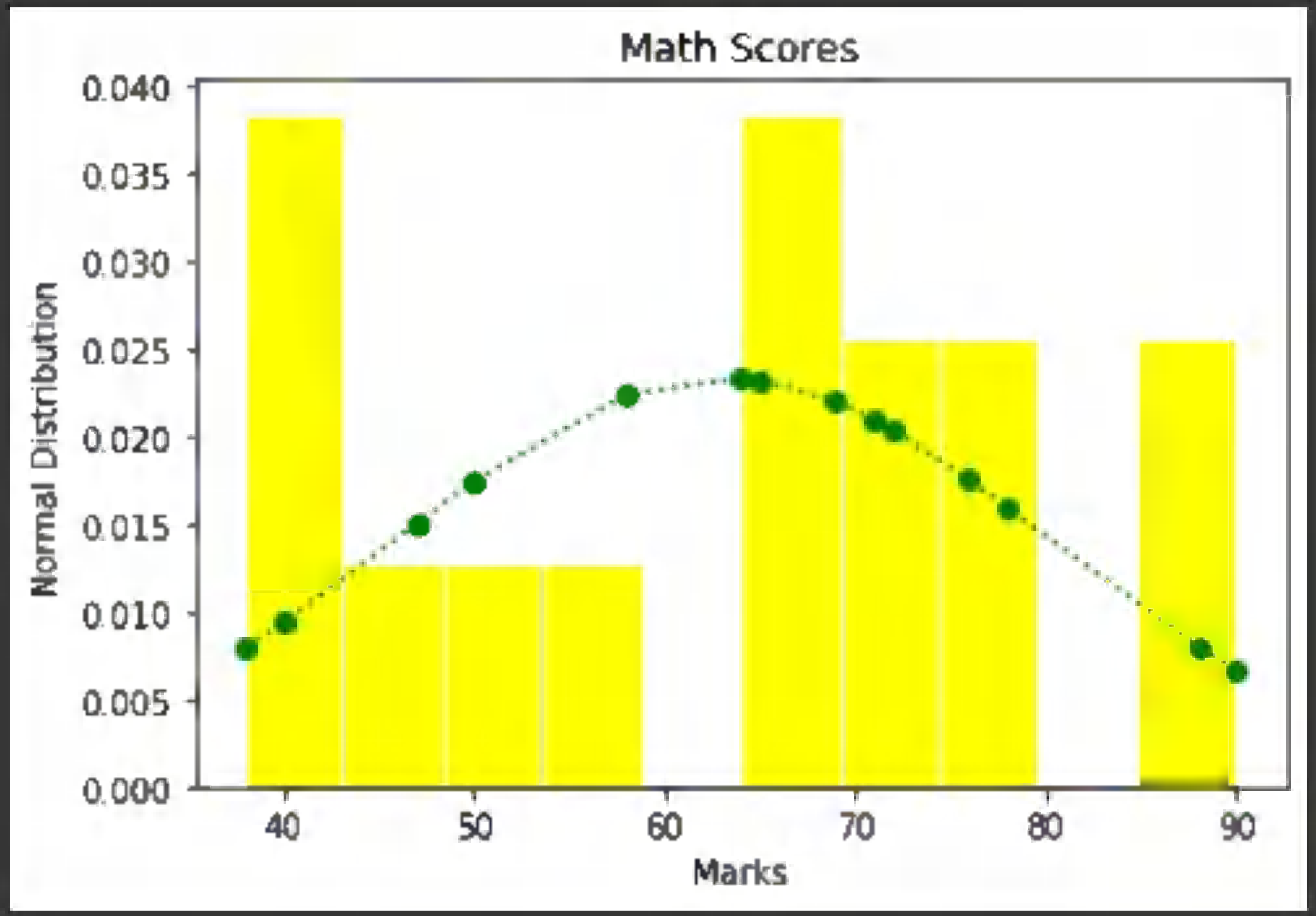
+ Code + Text Connecting Editing

Files

sample_data

```
[ ] mean= np.mean(x)
    sd = np.std(x,ddof =1)
    median = np.median(x)
    fit = stats.norm.pdf(x,mean,sd)
```

```
[ ] plt.hist(x, density = True, color ="yellow",ec ="white")
    plt.plot(x, fit,"go:")
    plt.title("Math Scores")
    plt.xlabel("Marks")
    plt.ylabel("Normal Distribution")
    plt.show()
```

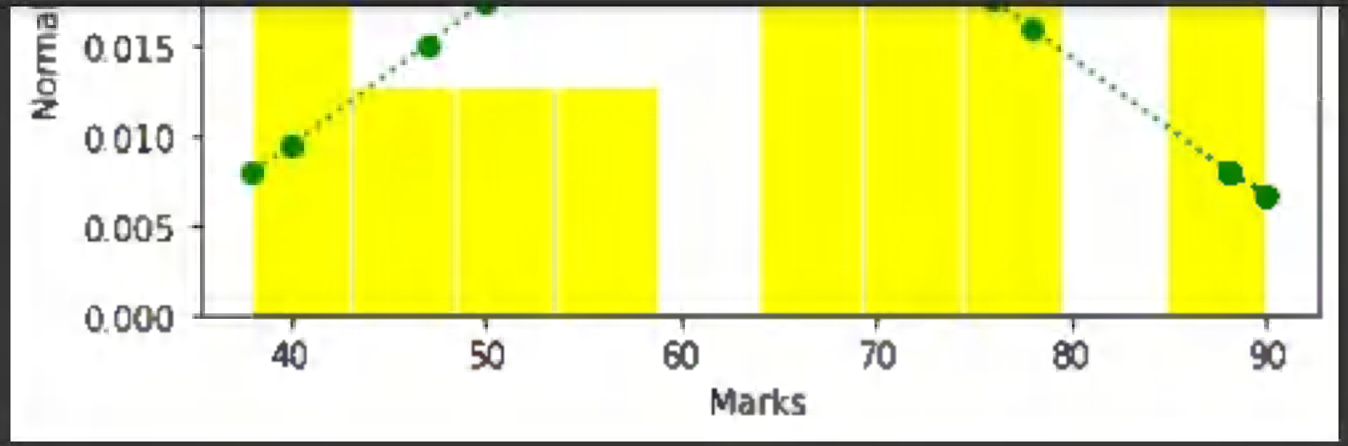


Skewness

Type here to search

24°C Light rain ENG 13:55 13-01-2022

[ ]



Skewness

defination:

1)Skewness measures the shift of the distribution from the normal bell curve

2)Positive skew value denotes right shift whereas negative skew value denotes left shift.

```
[ ] def skewness(x,mean,sd):
        return np.sum((x-mean)**3)/((len(x)-1)*(sd**3))
```

```
[ ] skewness(x,mean,sd)
```

```
-0.06988800386918068
```

Kurtosis

Defination

1)Kurtosis define how thick the tail, ends of distribution, which gives the probability of finding extremes

AIML Assignment2.ipynb

```
    return np.sum((x-mean)**3)/((len(x)-1)*(sd**3))
```

```
skewness(x,mean,sd)
```

```
-0.06988800386918068
```

Kurtosis

Defination

1)Kurtosis define how thick the tail- ends of distribution, which gives the probability of finding extremes.

2)It is the fourth central movement.

```
def kurtosis(x,mean,sd):
    return np.sum((x-mean)**4)/((len(x)-1)*(sd**4))
```

```
kurtosis(x,mean,sd)
```

```
1.7431786122403565
```