Stats Assignment 1

Calculate the mean, median, mode and standard deviation for the problem statements 1& 2.

Problem Statement 1:

The marks awarded for an assignment set for a Year 8 class of 20 students were as follows:

6 7 5 7 7 8 7 6 9 7 4 10 6 8 8 9 5 6 4 8

Answer:-

Input:-

```
import pandas as pd
l=[6 ,7 ,5 ,7 ,7 ,8 ,7 ,6 ,9 ,7 ,4 ,10 ,6 ,8 ,8 ,9 ,5 ,6,4,8]
a=pd.Series(l)
# calculate Mean, Median, mode, standard deviation.
print("Mean is:-",a.mean())
print("Median is:-",a.median())
print("Mode is:-",a.mode())
print("Standard Deviation is:-",a.std())
```

Output:-

```
Mean is:- 6.85
Median is:- 7.0
Mode is:- 0    7
dtype: int64
Standard Deviation is:- 1.6311119875071343
```

Problem Statement 2:

The number of calls from motorists per day for roadside service was recorded for a particular month:

28, 122, 217, 130, 120, 86, 80, 90, 140, 120, 70, 40, 145, 113, 90, 68, 174, 194, 170, 100, 75, 104, 97,75,123, 100, 75, 104, 97, 75, 123, 100, 89, 120, 109

Answer:-

Input:-

```
l1=[28, 122, 217, 130, 120, 86, 80, 90, 140, 120, 70, 40, 145, 113, 90, 68, 174, 194, 170,100, 75, 104, 97,
 75,123, 100, 75, 104, 97, 75, 123, 100, 89, 120, 109]
b=pd.Series(l1)
b
# Calculate Mean, Median, mode, standard deviation.
print("Mean is:-",b.mean())
print("Median is:-",b.median())
print("Mode is:-",b.mode())
print("Standard Deviation is:-",b.std())
```

Output:-

```
Mean is:- 107.51428571428572
Median is:- 100.0
Mode is:- 0    75
dtype: int64
Standart Deviation is:- 39.33892805484411
```

Problem Statement 3:

The number of times I go to the gym in weekdays, are given below along with its associated probability: x = 0, 1, 2, 3, 4, 5

f(x) = 0.09, 0.15, 0.40, 0.25, 0.10, 0.01

Calculate the mean no. of workouts in a week. Also evaluate the variance involved in it.

Answer:-

Input:-

```
x=[0,1,2,3,4,5]
fx=[0.09, 0.15, 0.40, 0.25, 0.10, 0.01]
meanofx=0
for i in x:
   meanofx+=x[i]*fx[i]
print("Means :-",meanofx)
varx=0
for i in x:
  varx=((x[i]-meanofx)**2)*fx[i]
print("Variance :-",varx)
```

output:-

```
Means :- 2.15
Variance :- 0.081225
```

Problem Statement 4:

Let the continuous random variable D denote the diameter of the hole drilled in an aluminium sheet. The target diameter to be achieved is 12.5mm. Random disturbances in the process often result in inaccuracy.

Historical data shows that the distribution of D can be modelled by the PDF (d) = 20e−20(d−12.5), d ≥ 12.5. If a part with diameter > 12.6 mm needs to be scrapped, what is the proportion of those parts? What is the CDF when the diameter is of 11 mm? What is your conclusion regarding the proportion of scraps?

Answer:-

For the drilling operation:-

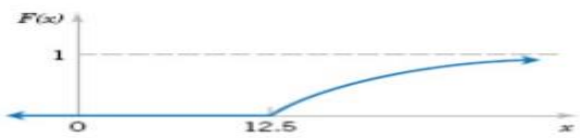
$$F(x) = 0 \quad \text{for} \quad x < 12.5$$

and for $12.5 \leq x$

$$F(x) = \int_{12.5}^{x} 20e^{-20(u-12.5)}\,du = 1 - e^{-20(x-12.5)}$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 12.5 \\ 1 - e^{-20(x-12.5)} & 12.5 \leq x \end{cases}$$



F(x) consists of two operations,

$$F(x) = 0 \quad \text{for} \quad x < 12.5$$

$$F(x) = \int_{12.5}^{x} 20e^{-20(u-12.5)}\,du$$
$$= 1 - e^{-20(x-12.5)} \qquad \text{for } 12.5 \le x$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 12.5 \\ 1 - e^{-20(x-12.5)} & 12.5 \le x \end{cases}$$

Problem Statement 5:

A company manufactures LED bulbs with a faulty rate of 30%. If I randomly select 6 chosen LEDs, what is the probability of having 2 faulty LEDs in my sample?

Calculate the average value of this process. Also evaluate the standard deviation associated with it.

Answer:-

Input:-

```
n = 6
x = 2
p = 0.3
print(stats.binom.pmf(x,n,p))
print('Average value =',n*p)
print('Standar Deviation =',np.sqrt(n*p*(1-p)))
```

Output:-

```
0.32413499999999995
Average value = 1.7999999999999998
Standar Deviation = 1.1224972160321822
```

Problem Statement 6:

Gaurav and Barakha are both preparing for entrance exams. Gaurav attempts to solve 8 questions per day with a correction rate of 75%, while Barakha averages around 12 questions per day with a correction rate of 45%. What is the probability that each of them will solve 5 questions correctly? What happens in cases of 4 and 6 correct solutions? What do you infer from it? What are the two main governing factors affecting their ability to solve questions correctly? Give a pictorial representation of the same to validate your answer.

Answer:-

Input:-

```
lambdag = 8*0.75
lambdab = 12*.45
g = stats.poisson.pmf(5,lambdag)
b = stats.poisson.pmf(5,lambdab)
g1 = stats.poisson.pmf(4,lambdag)
b1 = stats.poisson.pmf(6,lambdab)
print(round(g*100,2),'% probabilty that Gaurav will solve 5 questions correctly')
print(round(b*100,2),'% probabilty that Barakha will solve 5 questions correctly')
print(round(g1*100,2),'% probabilty that Gaurav will solve 4 questions correctly')
print(round(b1*100,2),'% probabilty that Barakha will solve 6 questions correctly')
```

Output:-

```
16.06 % probabilty that Gaurav will solve 5 questions correctly
17.28 % probabilty that Barakha will solve 5 questions correctly
13.39 % probabilty that Gaurav will solve 4 questions correctly
15.55 % probabilty that Barakha will solve 6 questions correctly
```

So we can infer that , the governing factor here is accuracy.

Problem Statement 7:

Customers arrive at a rate of 72 per hour to my shop. What is the probability of k customers arriving in 4 minutes? a) 5 customers, b) not more than 3 customers, c) more than 3 customers. Give a pictorial representation of the same to validate your

Answer:-

Input:-

```
lambda_rate = 72 * 4 / 60
a = stats.poisson.pmf(5,lambda_rate)
b = stats.poisson.cdf(2,lambda_rate)
c = stats.poisson.sf(2,lambda_rate)
print('Probability that 5 customers arriving in 4 minutes is :',a)
print('Probability that not more than 3 customers arriving in 4 minutes is :',b)
print('Probability that more than 3 customers arriving in 4 minutes is :',c)
```

Output:-

```
Probability that 5 customers arriving in 4 minutes is: 0.17474768364388296
Probability that not more than 3 customers arriving in 4 minutes is: 0.1425392188890269
Probability that more than 3 customers arriving in 4 minutes is: 0.8574607811109731
```

Problem Statement 8:

I work as a data analyst in Aeon Learning Pvt. Ltd. After analyzing data, I make reports, where I have the efficiency of entering 77 words per minute with 6 errors per hour. What is the probability that I will commit 2 errors in a 455-word financial report? What happens when the no. of words increases/decreases (in case of 1000 words, 255 words)?

How is the λ affected?

How does it influence the PMF?

Give a pictorial representation of the same to validate your answer.

Answer:-

Input:-

```
error1 = 455 * 6 / 77
print('the probability that I will commit 2 errors in a 455-word financial report
is',stats.poisson.pmf(2,error1))
error2 = 1000 * 6 / 77
print('the probability that I will commit 2 errors in a 455-word financial report
is',stats.poisson.pmf(2,error2))
error3 = 255 * 6 / 77
print('the probability that I will commit 2 errors in a 455-word financial report
is',stats.poisson.pmf(2,error3))
```

Output:-

```
the probability that I will commit 2 errors in a 455-word financial report
is 2.51536143880336e-13
the probability that I will commit 2 errors in a 455-word financial report
is 4.3768593897408885e-31
the probability that I will commit 2 errors in a 455-word financial report
is 4.633227851444488e-07
```

So when the words increase , the probabilty of making 2 errors decreases.

Problem Statement 9:

Let the continuous random variable D denote the diameter of the hole drilled in an aluminium sheet. The target diameter to be achieved is 12.5mm. Random disturbances in the process often result in inaccuracy. Historical data shows that the distribution of D can be modelled by the PDF, f(d) =

20e−20(d−12.5), d ≥ 12.5. If a part with diameter > 12.6 mm needs to be scrapped, what is the proportion of those parts? What is the CDF when the diameter is of 11 mm? What is the conclusion of this experiment?

Answer:-

For the drilling operation:-

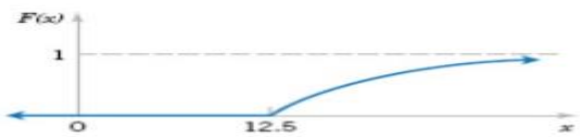
$$F(x) = 0 \quad \text{for} \quad x < 12.5$$

and for $12.5 \leq x$

$$F(x) = \int_{12.5}^{x} 20e^{-20(u-12.5)}\,du = 1 - e^{-20(x-12.5)}$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 12.5 \\ 1 - e^{-20(x-12.5)} & 12.5 \leq x \end{cases}$$



F(x) consists of two operations,

$$F(x) = 0 \quad \text{for} \quad x < 12.5$$

$$F(x) = \int_{12.5}^{x} 20e^{-20(u-12.5)}\, du$$

$$= 1 - e^{-20(x-12.5)} \qquad \text{for } 12.5 \le x$$

Therefore,

$$F(x) = \begin{cases} 0 & x < 12.5 \\ 1 - e^{-20(x-12.5)} & 12.5 \le x \end{cases}$$

Problem Statement 10:

Please compute the following:

a) P(Z > 1.26), . P(Z < −0.86), P(Z > −1.37), P(−1.25 < Z < 0.37), . P(Z ≤ −4.6)

b) Find the value z such that P(Z > z) = 0.05

c) Find the value of z such that P(−z < Z < z) = 0.99

Answer:-

a) P (Z > 1.26) = 0.10384
   P (Z < −0.86) = 0.19490
   P (Z > −1.37) = 0.91465
   P (−1.25 < Z < 0.37) = 0.53866
   P (Z ≤ −4.6) = Nearly Zero

b) Z = 1.65

c) Z = 2.58

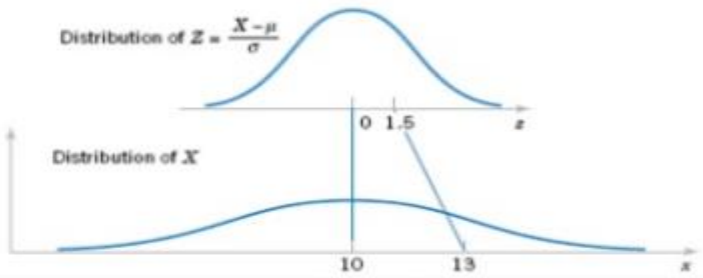
Problem Statement 11:

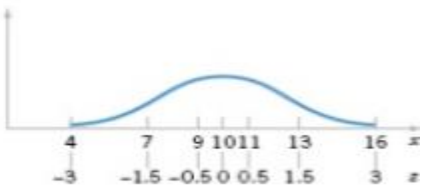
The current flow in a copper wire follow a normal distribution with a mean of 10 mA and a variance of 4 $(mA)^2$. What is the probability that a current measurement will exceed 13 mA? What is the probability that a current measurement is between 9 and 11mA? Determine the current measurement which has a probability of 0.98.

Answer:-

P(X>13) = P((x-10)/2>(13-10)/2)   =p(Z>1.5)   =0.06681

$P(X>13) = P(Z>1.5) = 1-P(Z \le 1.5) = 1 - 0.93319 = 0.06681$





$P(9<X<11) = P((9-10)/2<(X-10)/2<(11-10)/2$

$= P(-0.5<Z<0.5) \quad = P(Z<0.5)-P(Z< -0.5)$

$= 0.69146 – 0.30854 = 0.38292$

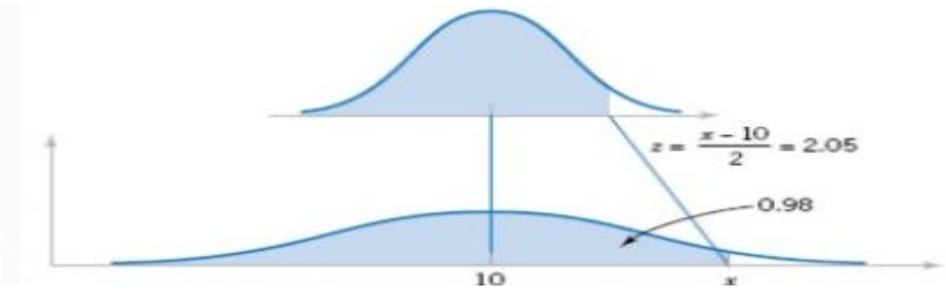
$P(X<x) = P((X-10)/2<(x-10/2)$

$= P(Z<(x-10)/2)$

$=0.98$

$P(Z<2.05)=0.97982$

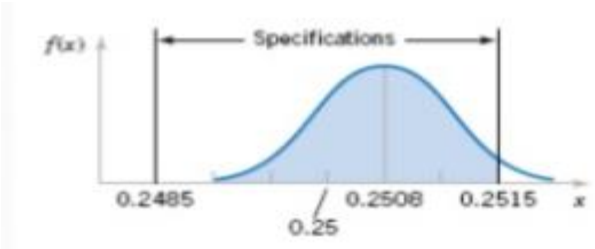
$x=2(2.05)+10$ = 14.1 miliamperes



Problem Statement 12:

The shaft in a piston has its diameter normally distributed with a mean of 0.2508 inch and a standard deviation of 0.0005 inch. The specifications of the shaft are 0.2500 ∓ 0.0015 inch. What proportion of shafts is in sync with the specifications? If the process is centered so that the mean is equal to the target value of 0.2500, what proportion of shafts conforms to the new specifications? What is your conclusion from this experiment?

Answer:-



Let X denotes the shaft diameter in inches.

P(0.2485 < X < 0.2515) = P((0.2485-0.2508/0.0005)<Z<(0.2515-0.2508)/0.0005)

=P(-4.6 <Z <1.4) = P(Z<1.4)-P(Z< -4.6)

=0.91924-0.0000=0.91924