Problem statement1:

Calculate the mean, median, mode and standard deviation for the problem statements 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

Solution:

In [55]:
```
#Problem statement1
import numpy as np
import statistics
l = [6,7,5,7,7,8,7,6,9,7,4,10,6,8,8,9,5,6,4,8]
total = sum(l)
mean = total/len(l)
print("mean: ", mean)
median = np.median(l)
print("median: ",median)
mode = statistics.multimode(l)
print("mode: ", mode)
standard_deviation = statistics.pstdev(l)
print("Standard_deviation: ",standard_deviation)
```

```
mean:  6.85
median:  7.0
mode:  [7]
Standard_deviation:  1.5898113095584647
```

Problem Statement 2:

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

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

Solution:

```
#Problem statement2
import numpy as np
import statistics
l = [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
total = sum(l)
mean = total/len(l)
print("mean: ", mean)
median = np.median(l)
print("median: ",median)
mode = statistics.multimode(l)
print("mode: ", mode)
standard_deviation = statistics.pstdev(l)
print("Standard_deviation: ",standard_deviation)
```

```
mean:  107.51428571428572
median:  100.0
mode:  [75]
Standard_deviation:  38.77287080168403
```

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

Solution:.

Formula of mean is

$$E(X) = \mu = \sum xP(x).$$

Formula of variance is

$$\sigma^2 = \sum (x - \mu)^2 P(x) .$$

```
#Problem statement3
r = [0,1,2,3,4,5]
p = [0.09,0.15,0.40,0.25,0.10,0.01]
mean = 0
for (i,j) in zip(r,p):
     mean = mean + (i * j)
print("mean :",mean)
variance = 0
for (i,j) in zip(r,p):
    variance = variance + (((i-mean)**2)*j)
print("Variance :",variance)
```

```
mean : 2.15
Variance : 1.2275
```

Problem Statement 4:
Let the continuous random variable D denote the diameter of the hole drilled in an aluminum 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?
Solution:

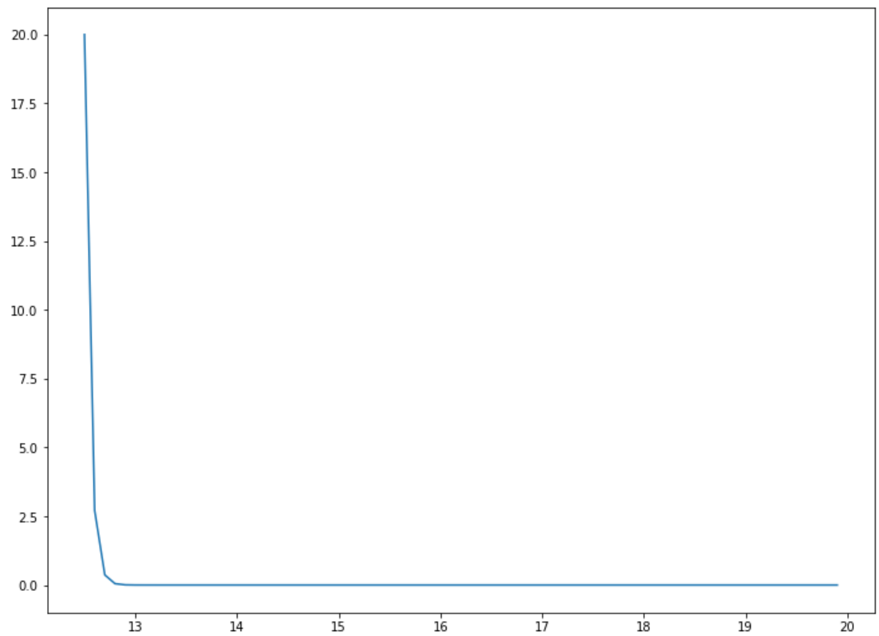
```
In [30]: #Problem 9
from scipy import integrate
from scipy.integrate import quad
import numpy as np
def integrand(x):
    return 20*np.exp(-20*(x-12.5))
cdf_a = quad(integrand,12.6,np.inf)
print("Proportion of parts when scrapped at a diameter greater than 12.6 is : ",round(cdf_a[0],3))

Proportion of parts when scrapped at a diameter greater than 12.6 is :  0.135
```

Since PDF is defined is defined for d>=12.5, cdf when d= 11 is zero
Conclusion:
For diameter less than 12.5, the CDF will be zero
CDF equation for diameter greater than or equal to 12.5 is given by

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

We can plot this:

```
]: import matplotlib.pyplot as plt
x = np.arange(12.5,20,0.1)
def pdf_temp(x):
    pdf_values=[]
    for i in x:
        pdf_values.append(20*np.exp(-20*(i-12.5)))
    return pdf_values
fig = plt.figure(figsize=(12,9))
plt.plot(x,pdf_temp(x))
plt.show()
```



As diameter increases, proportion reduces
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.

Solution:

Formula to find probability in binomial distribution:

$$P(X) = C^n_x p^x q^{n-x}$$

Formula of mean is:

Mean = n*p

Standard deviation = sqrt[n * P * (1 - P)]

```
In [16]: #Problem statement5
from scipy.stats import binom
n = 6
r = 2
p = 0.3
q = 1-0.3
dist = binom.pmf(2,6,0.3)
print("probability of having 2 faulty LEDs in the sample is: ",dist)
mean =binom.stats(n,p,moments='m')
variance =binom.stats(n,p,moments='v')
print("mean is: ",mean)
print("standard deviation is: ",variance**0.5)
```

```
probability of having 2 faulty LEDs in the sample is:  0.32413499999999995
mean is:  1.7999999999999998
standard deviation is:  1.1224972160321822
```

Problem Statement 6:

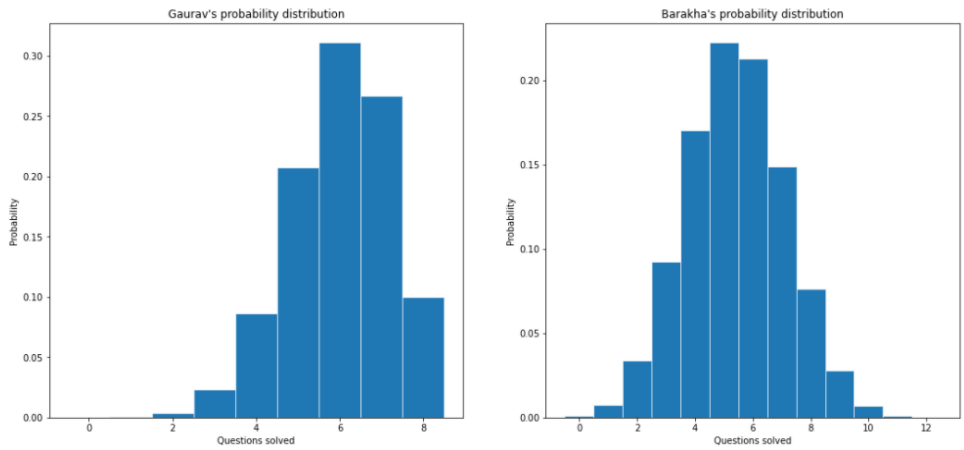
Gaurav and Barakha are both preparing for entrance exams. Gaurav attempts to solve 8 questions per day with a correction rate of 75%, while Barakha averages around 12 questions per day with a correction rate of 45%. What is the probability that each of them will solve 5 questions correctly? What happens 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.

Solution:

Formula to find probability is: $P(X) = C^n_x p^x q^{n-x}$

In [7]:
```python
#Problem statement 6
#Probability of Gaurav attending 5,4,6 question correctly
from scipy.stats import binom
n =8
r =5
p_G = 0.75
dist = binom.pmf(r,n,p_G)
print("Probability of Gaurav solves 5 question correctly is: ", dist)
r = 4
dist = binom.pmf(r,n,p_G)
print("Probability of Gaurav solves 4 question correctly is: ", dist)
r = 6
dist = binom.pmf(r,n,p_G)
print("Probability of Gaurav solves 6 question correctly is: ", dist)
#Probability of Barakha attending 5,4,6 question correctly
m =12
r =5
p_B = 0.45
dist = binom.pmf(r,m,p_B)
print("Probability of Barakha solves 5 question correctly is: ", dist)
r = 4
dist = binom.pmf(r,m,p_B)
print("Probability of Barakha solves 4 question correctly is: ", dist)
r = 6
dist = binom.pmf(r,m,p_B)
print("Probability of Barakha solves 6 question correctly is: ", dist)
Gaurav_prob_values = [binom.pmf(x,n,p_G) for x in list(range(n+1))]
Barakha_prob_Values = [binom.pmf(x,m,p_B) for x in list(range(m+1))]
Gaurav_correct_answers = list(range(n+1))
Barakha_correct_answers = list(list(range(m+1)))
import matplotlib.pyplot as plt
fig = plt.figure(figsize=(18,8))
plt.subplot(1, 2, 1)
plt.title("Gaurav's probability distribution")
plt.xlabel("Questions solved")
plt.ylabel("Probability")
plt.bar(Gaurav_correct_answers,Gaurav_prob_values, width=1, edgecolor="white", linewidth=0.7)
plt.subplot(1, 2, 2)
plt.title("Barakha's probability distribution")
plt.xlabel("Questions solved")
plt.ylabel("Probability")
plt.bar(Barakha_correct_answers,Barakha_prob_Values, width=1, edgecolor="white", linewidth=0.7)
plt.show()
```

```
Probability of Gaurav solves 5 question correctly is:  0.20764160156250022
Probability of Gaurav solves 4 question correctly is:  0.08651733398437506
Probability of Gaurav solves 6 question correctly is:  0.31146240234375017
Probability of Barakha solves 5 question correctly is:  0.22249823843265792
Probability of Barakha solves 4 question correctly is:  0.16996393213605795
Probability of Barakha solves 6 question correctly is:  0.21238468214026424
```



Conclusion:
As the rate of correction increases, the skewness also increases.
As the no of questions solved increases, the skewness decreases

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.

Solution:

Formula for this is Poisson's formula.

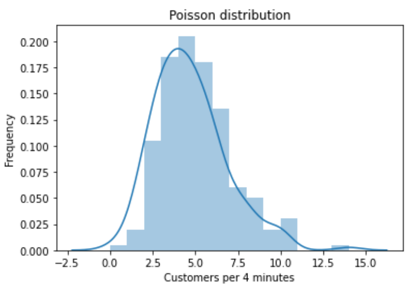
$$P(x; \mu) = (e^{-\mu}) (\mu^x) / x!$$

```
[7]: #Problem statement7
from scipy.stats import poisson
mu = (72/60)*4
probability_a = poisson.pmf(5,mu)
print("Probability of 5 customers :",probability_a)
probability_b = poisson.pmf(0,mu) + poisson.pmf(1,mu)+ poisson.pmf(2,mu) + poisson.pmf(3,mu)
print("Probability of not more than 3 customers :",probability_b)
probability_c = 1 - probability_b
print("Probability of more than 3 customers :",probability_c)
```

```
Probability of 5 customers : 0.17474768364388296
Probability of not more than 3 customers : 0.29422991649656405
Probability of more than 3 customers : 0.705770083503436
```

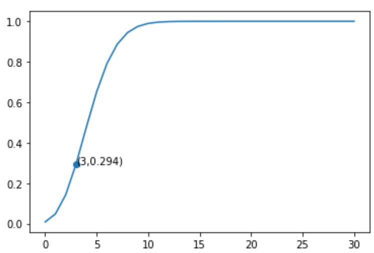
```
In [3]: import seaborn as sb
import pandas as pd
import matplotlib.pyplot as plt
data_poisson = poisson.rvs(4.8,size=200)
ax = sb.distplot(data_poisson)
ax.set(xlabel='Customers per 4 minutes', ylabel='Frequency')
plt.title('Poisson distribution')
plt.show()
```

```
C:\Users\91984\anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated function an
d will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar fle
xibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```




```
In [38]: x = list(range(31))
p_values = poisson.cdf(x,4.8)
y= p_values[3]
print('Probability of not more than 3 customers is',round(y,3))
print('Probability of more than 3 customers is',round((1 - y),3))
plt.plot(p_values)
plt.scatter(3,p_values[3])
plt.text(3,p_values[3],'({},{})'.format(3,round(p_values[3],3)))
plt.show()
```

```
Probability of not more than 3 customers is 0.294
Probability of more than 3 customers is 0.706
```



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 lamda affected?

How does it influence the PMF?

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

Solution:

Formula is

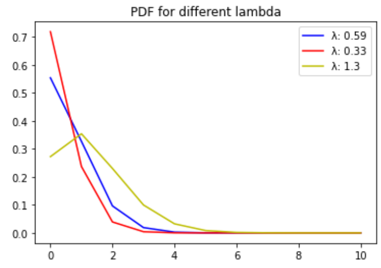
$$P(x; \mu) = (e^{-\mu}) (\mu^{x}) / x!$$

In [31]:

```
#Problem 8
from scipy.stats import poisson
words_per_minute = 77
Mean_0 = round(6/60,2)
print("Average error in a minute :",Mean_0)
Time_a = round(455/77,1)
print("Time taken to type 455 words :",Time_a)
Mean_a = round(Mean_0 * Time_a,2)
print("Average error in {} minutes is :  {}".format(Time_a,Mean_a))
prob_a =  round(poisson.pmf(2,Mean_a),5)
print("probability that I will commit 2 errors in a 455-word financial report is :",prob_a)
print("*"*50)
Time_b = round(255/77,1)
print("Time taken to type 255 words :",Time_b)
Mean_b = round(Mean_0 * Time_b,2)
print("Average error in {} minutes is :  {}".format(Time_b,Mean_b))
prob_b =  round(poisson.pmf(2,Mean_b),5)
print("probability that I will commit 2 errors in a 255-word financial report is :",prob_b)
print("*"*50)
Time_c = round(1000/77,1)
print("Time taken to type 1000 words :",Time_c)
Mean_c = round(Mean_0 * Time_c,2)
print("Average error in {} minutes is :  {}".format(Time_c,Mean_c))
prob_c =  round(poisson.pmf(2,Mean_c),5)
print("probability that I will commit 2 errors in a 1000-word financial report is :",prob_c)
```

```
Average error in a minute : 0.1
Time taken to type 455 words : 5.9
Average error in 5.9 minutes is :  0.59
probability that I will commit 2 errors in a 455-word financial report is : 0.09648
**************************************************
Time taken to type 255 words : 3.3
Average error in 3.3 minutes is :  0.33
probability that I will commit 2 errors in a 255-word financial report is : 0.03915
**************************************************
Time taken to type 1000 words : 13.0
Average error in 13.0 minutes is :  1.3
probability that I will commit 2 errors in a 1000-word financial report is : 0.23029
```

In [35]:
```python
#Checking how Lamda is varying
#λ: 0.59
import matplotlib.pyplot as plt
x = list(range(11))
plt.plot(x,poisson.pmf(x,Mean_a),'b-',label='λ: 0.59')
plt.plot(x,poisson.pmf(x,Mean_b),'r-',label='λ: 0.33')
plt.plot(x,poisson.pmf(x,Mean_c),'y-',label='λ: 1.3')
plt.legend()
plt.title('PDF for different lambda')
plt.show()
```



Conclusion
when lambda increases, data peak decreases. Also data spread is increasing with increasing lambda

Problem Statement 9:

Let the continuous random variable D denote the diameter of the hole drilled in an
aluminum 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?

Solution:

In [30]:
```python
#Problem 9
from scipy import integrate
from scipy.integrate import quad
import numpy as np
def integrand(x):
    return 20*np.exp(-20*(x-12.5))
cdf_a = quad(integrand,12.6,np.inf)
print("Proportion of parts when scrapped at a diameter greater than 12.6 is : ",round(cdf_a[0],3))
```

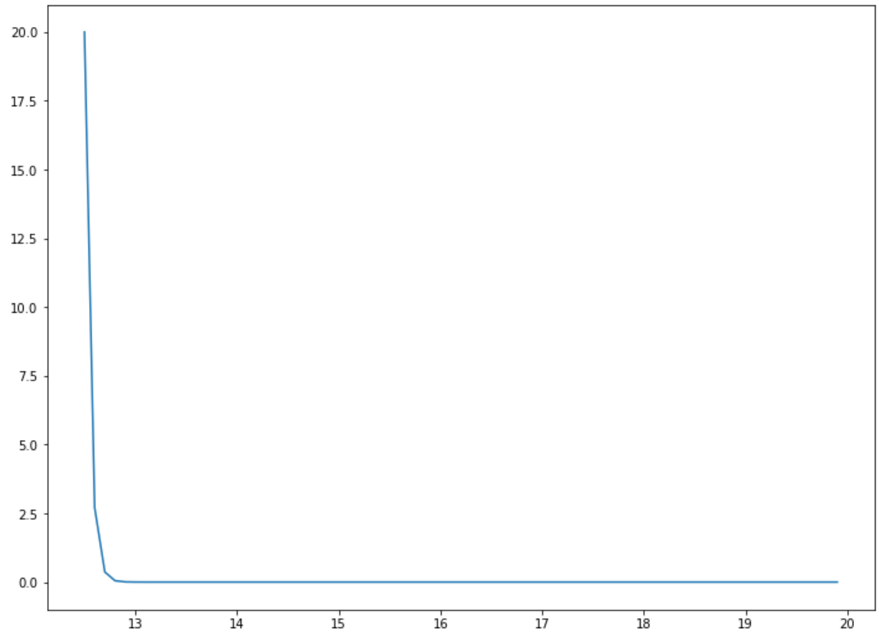
```
Proportion of parts when scrapped at a diameter greater than 12.6 is :  0.135
```

Since PDF is defined is defined for d>=12.5, cdf when d= 11 is zero
Conclusion:
For diameter less than 12.5, the CDF will be zero
CDF equation for diameter greater than or equal to 12.5 is given by

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

We can plot this:

```
]: import matplotlib.pyplot as plt
x = np.arange(12.5,20,0.1)
def pdf_temp(x):
    pdf_values=[]
    for i in x:
        pdf_values.append(20*np.exp(-20*(i-12.5)))
    return pdf_values
fig = plt.figure(figsize=(12,9))
plt.plot(x,pdf_temp(x))
plt.show()
```



As diameter increases, proportion reduces

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 z of such that p(−z< Z<z) = 0.99

Solution:

a) P(Z > 1.26) = 1-P(Z<1.26) = 0.10383

P(Z<-0.86) = 0.19489

P(Z>-1.37) = 1 - P(Z<-1.37) = 0.1466

P(-1.25<Z<0.37) = P(Z<-0.37)) - P(Z<-1.25) = 0.53866

P(Z≤4.6) = 0

b) P(Z>z) = 0.05

Then 1-0.05 = 0.95

So from Z table z= 1.64

c) So 1-0.99 = 0.01

So 0.01/2 = 0.005

P(Z<z) = 0.005

So z = 2.58 0r -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 9ma and 11mA? Determine the
current measurement which has a probability of 0.98.

Solution:
Mean = 10ma
Standard deviation = spqrt(4) = 2
X= 13
Z = (X- Standard deviation)/Standard deviation
= (13-10)/2
= 1.5
Probability that current measurement exceed 13ma = P(Z>1.5) = 1-P(Z<1.5) = 1-0.093319 = 0.06681

Probability that a current measurement is between 9ma and 11mA:
X=9
Z score of 9 = (9-10)/2 = -0.5
Z score of 11 = (11-10)/2 = 0.5

So Probability that a current measurement is between 9ma and 11mA = P(Z<0.5) - P(Z<-0.5)= 0.691 - 0.309 = 0.382
current measurement which has a probability of 0.98 :
From Z table P(Z<z) = 0.98,
Then z approximately 2.05
X = (Z * Standard deviation) + Mean
= 2.05*2+10
= 14.1mA

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 are 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 conform to the new specifications? What is your conclusion from this experiment?

Solution:
When X = 0.2500-0.0015
Mean = 0.2508
$\sigma = 0.0005$
Z score = ((0.2500-0.0015) - 0.2508)/0.0005 = -4.6
When X = 0.2500+0.0015
Mean = 0.2508
$\sigma = 0.0005$
Z score = ((0.2500+0.0015) - 0.2508)/0.0005 = 1.4
Proportion of shafts = P(-4.6<Z<1.4) = P(Z<1.4) - P(Z<-4.6) = 0.91924-0.0002 = 0.92

When X = 0.2500-0.0015
Mean = 0.2500

$\sigma = 0.0005$
Z score = ((0.2500-0.0015) - 0.2500)/0.0005 = -3
When X = 0.2500+0.0015
Mean = 0.2500
$\sigma = 0.0005$
Z score = ((0.2500+0.0015) - 0.2500)/0.0005 = 3
Proportion of shafts = P(-3<Z<3) = P(Z<3) - P(Z<-3) = 0.99865-0.0135 = 0.98

Conclusion:
As mean decreases , proportion increases