statistics-lab-5-assignment-1

November 29, 2023

Q1. Solve using python code

a) A store sells 3 apples per day on average. What is the probability that they will sell 5 apples on a given day?

[9]:
```
from scipy.stats import poisson

lambda_val = 3
k = 5

probability = poisson.pmf(k, lambda_val)

print("The probability of selling exactly apples is: ", probability)
```

```
The probability of selling exactly apples is:  0.10081881344492458
```

b) A certain store sells seven footballs per day on average. What is the probability that this store sells four or less footballs in a given day?

[8]:
```
from scipy.stats import poisson

lambda_val = 7
k = 4

probability = poisson.cdf(k, lambda_val)

print("The probability of selling four or fewer footballs is: ",probability )
```

```
The probability of selling four or fewer footballs is:  0.17299160788207146
```

c) A certain store sells 15 cans of tuna per day on average. What is the probability that this store sells more than 20 cans of tuna in a given day?

[7]:
```
from scipy.stats import poisson

lambda_val = 15
k = 20

probability = 1 - poisson.cdf(k, lambda_val)
```

```
print("The probability of selling more than 20 cans of tuna is :", probability)
```

```
The probability of selling more than 20 cans of tuna is : 0.08297091003146029
```

Q2. Explain the question comes in which distribution and why?

a) If you walked up to a random person on the street, the probability that their birthday falls on a given date.

Ans= It is a uniform distribution, because it assumes that each day of the year is equally likely to be the birthday of a random person. This means that the probability of a person having a birthday on a given date is the same for all dates, and does not depend on any other factors. The uniform distribution is a type of probability distribution that assigns the same probability to any possible outcome within a specified range.If we pick a random person on the street, the probability of their birthday falling on a given date is 1/365, and this is also a uniform distribution.

b) Suppose it's known that 5% of all widgets on an assembly line are defective. De-termine the probability of inspecting 0, 1, 2 widgets, etc. before an inspector comes across a defective widget.

Ans= This is a problem that can be solved using the geometric distribution. The geometric distribution describes the probability of experiencing a certain amount of failures before experiencing the first success. In this case, the success is finding a defective widget, and the failure is finding a non- defective widget. The probability of success is p = 0.05, and the probability of failure is q = 1 - p = 0.95. The formula for the geometric distribution is: Pr(X = k) = q^(k-1) * p where X is the number of trials until the first success, and k is a positive integer. Using this formula, we can calculate the probability of inspecting 0, 1, 2 widgets, etc. before finding a defective widget:

Pr(X = 1) = q^(1-1) * p = 0.95^0 * 0.05 = 0.05 Pr(X = 2) = q^(2-1) * p = 0.95^1 * 0.05 = 0.0475 Pr(X = 3) = q^(3-1) * p = 0.95^2 * 0.05 = 0.0451 Pr(X = 4) = q^(4-1) * p = 0.95^3 * 0.05 = 0.0428 Pr(X = 5) = q^(5-1) * p = 0.95^4 * 0.05 = 0.0406

and so on. The expected value (mean) of the geometric distribution is E(X) = 1/p = 1/0.05 = 20, which means that on average, the inspector will have to inspect 20 widgets before finding a defective one.

c) Find the probability that a certain number of patients will experience side effects as a result of taking new medications.

Ans= Medical professionals use the binomial distribution to model the probability that a certain number of patients will experience side effects as a result of taking new medications.

For example, suppose it is known that 5% of adults who take a certain medication experiencenegative side effects. We can u se a Binomial Distribution Calculator to find the probability that more than a certain number of patients in a random sample of 100 will experience negative side effects.

P(X > 5 patients experience side effects) = 0.38400 P(X > 10 patients experience side effects) = 0.01147 P(X > 15 patients experience side effects) = 0.0004 And so on.

This gives medical professionals an idea of how likely it is that more than a certain number of patients will experience negative side effects.

Q3. Explain different sampling methods and show it using python code.

1)Random Sampling.

[11]:
```
import numpy as np

population = np.arange(1,101)
random_sample = np.random.choice(population, size=15, replace=False)

print("Population : ", population)
print("*"*100)
print("Random Sample : ", random_sample)
```

```
Population :  [  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16
17  18
  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72
  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
  91  92  93  94  95  96  97  98  99 100]
****************************************************************************
********************
Random Sample :  [89 97 66 88 34 18 68 39 57 37 94 29 86  6  3]
```

2) Stratified Sampling.

[12]:
```
import numpy as np

strata1 = np.random.normal(40,7,100)
strata2 = np.random.normal(80,14,200)

sample_strata1 = np.random.choice(strata1, size=10,replace=False)
sample_strata2 = np.random.choice(strata2, size=10,replace=False)

print("Strata1 : ", strata1)
print("*"*100)
print("Sample Strata 1 : ",sample_strata1)
print("*"*100)
print("*"*100)
print("Strata2 : ", strata2)
print("*"*100)
print("Sample Strata 2 : ",sample_strata2)
```

```
Strata1 :  [31.58696555 51.33836889 40.47362041 52.90310145 36.60224717
39.14322924
 43.83087487 41.780673   44.41529928 38.0451066  38.61350248 48.20509169
 35.96917441 34.98713033 37.33511352 34.67857041 29.84719542 37.63512396
 33.98069975 42.77995177 30.96644192 33.29967576 50.89981013 37.44304562
 46.60154891 23.75196677 33.65490794 37.66382421 44.00994935 48.98521809
 35.69383343 29.68334952 44.99530391 46.91815288 38.05653693 48.75591925
 37.29774107 42.64334776 34.15296855 33.3367863  49.11656663 47.01120695
 39.98588796 32.99257237 30.08298672 53.35021484 54.37678293 43.18611982
```

```
 31.78746562 46.10970364 34.80959789 34.67898614 45.27093857 56.84483733
 44.65378987 43.35304687 32.36903868 62.93944373 50.33533432 28.07343001
 39.50461695 33.33543439 45.95972909 43.07641801 28.54278239 43.86330224
 38.48974893 30.23238576 43.94002131 50.56725429 44.00238933 39.33017434
 34.61389902 40.58714239 35.87831013 42.54662819 39.97531114 47.20249078
 35.72524049 37.42904712 39.23276314 44.10397466 46.35235793 33.23663968
 48.44910052 44.90403022 46.12731129 48.77376523 48.53333712 38.66417064
 46.08993916 38.03345803 38.07627727 37.25218379 41.73118085 39.59821646
 44.53773272 38.2504716  33.32906221 41.96820023]
**************************************************************************
********************
Sample Strata 1 :  [43.86330224 38.66417064 48.75591925 54.37678293 34.15296855
46.08993916
 33.65490794 53.35021484 49.11656663 45.95972909]
**************************************************************************
********************
**************************************************************************
********************
Strata2 :  [ 83.6492605   45.97285135  89.7236105   81.61630409  95.49821068
  90.23051311  96.89380488  67.0348145   88.34273388  95.58686336
  79.32660243  95.43800974  95.90461012  95.06107714  92.66623157
  93.2614536   82.74413094  79.14052256 106.12323245 107.32196585
  82.82735986  74.1231602   73.71377384  73.63350456  91.67959075
 110.90351229  67.30296434  65.07691367  70.37063235  85.51053475
  73.44457242  89.89933956  77.39049263  92.79654159  90.38028768
  78.59159315  74.28024858  85.02979617  96.57353103  66.82252768
  87.76262645  61.12752532  68.94039357  85.00320237  68.21317914
  83.12108634  83.65767124  83.23251708  79.71332685  79.78460903
  72.0475867   88.37816739  74.50108932  65.62297077  82.83757709
  86.81912825 103.9361162   71.22034225  95.32164595  85.90860667
  85.27325801  85.50217558  94.31445706  63.19221514  66.23548715
  79.17066166 105.87164631  71.10012755  71.188645    65.9703566
  74.63933801  81.59494175  74.38926513  69.49395222  88.80461426
  66.67909067  70.7936251   84.37157579  64.92052131  77.89560827
  84.37932147  84.33952817  97.29628112  94.05684985 108.38940206
  66.60182826  57.47043982  73.32688492  88.58102643  99.04181032
 103.16567691  69.57446142  68.47053846  70.90099544 106.14220357
  96.16924339  96.19797169  86.33475978  72.50556027  76.87636714
  87.12795153  98.62224911  94.15516843  97.10756199  51.99423116
 103.78837059 103.58949314  85.44986452  65.59507788  84.50552607
  38.54026163  93.20812419 110.45677387  96.19710853  68.05336095
  97.52673593  88.04385412  85.61916758  92.42827666  78.88872197
  94.52992485  97.28665711  78.3251686   93.8392974   64.70098955
  92.29031223  74.33202422  58.3197129   63.58928334  70.48851921
  81.7527188   60.87970613  62.73929928  89.28348004  53.27839413
  50.37594505  74.55451716  93.67327089 102.34595843  80.97896239
 102.61950933 103.25341521  81.53329877  59.77813331  94.36211501
  82.36787633  81.79175252  62.53160865  83.33760027  61.69510285
```

```
  87.0062725   93.66677003  90.5226356   93.85102434  71.14448526
  75.94876898  75.81409063  65.84417344  76.32109457  90.91013262
  70.29317878  72.98898295  77.14135032  66.51602768  67.86869953
  82.69437733 100.98968556  60.88314792  67.24283124  89.76220352
  90.33103673  42.00865033  59.32487673  82.2136978   90.98173524
  61.18440676  68.96940177  65.93760651  76.03926788  59.61160288
  84.7663485   68.92158905  70.91532882  67.52261062  69.52907664
  59.28900763  78.44446283  73.49708165  62.9078582   77.33479602
  74.66280372  91.77424942  64.39989531  73.09300654  88.18234437
  93.10242997  70.14455906  88.2569202   81.81802467  69.62712783]
**********************************************************************
**********
Sample Strata 2 :  [ 81.59494175 106.12323245  70.7936251   83.33760027
93.67327089
  64.70098955  83.65767124  70.29317878  87.76262645  67.30296434]
```

3) Systematic Sampling.

[13]:
```
import numpy as np

population = np.arange(1,101)
k=10

systematic_sample = population[::k]

print("Population : ", population)
print("*"*100)
print("Systematic Sample : ", systematic_sample)
```

```
Population :  [  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16
17  18
  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72
  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
  91  92  93  94  95  96  97  98  99 100]
**********************************************************************
**********
Systematic Sample :  [ 1 11 21 31 41 51 61 71 81 91]
```

4) Cluster Sampling

[15]:
```
import numpy as np

clusters = [np.random.normal(i*20,5,40) for i in range(1,6)]

selected_cluster_indices = np.random.choice(len(clusters), size=2,␣
 ↪replace=False)
cluster_sample = np.concatenate([clusters[i] for i in selected_cluster_indices])
```

```
print(" Clusters : ", clusters)
print("*"*120)
print("Cluster Sample", cluster_sample)
```

```
 Clusters :  [array([26.00848614, 24.95270636, 20.57336732, 16.69351583,
30.1565901 ,
       22.93038686, 18.46838525, 18.59756875, 18.37035232, 20.81517623,
       12.35414791, 14.78214557, 22.40527504, 22.35443699, 16.88135277,
        8.73161621, 18.57594005, 14.98927277, 21.67312813, 19.53760705,
       20.03403219, 19.09243905, 14.41974289, 24.66501475, 17.93830184,
       14.24261113, 23.99777887, 20.66355947, 20.11985052, 19.04906346,
       19.35414147, 18.95591389, 26.33357363, 25.35612884, 30.62569754,
       22.21593111, 22.22606541, 21.8968823 , 17.67490954, 11.6816183 ]),
array([47.18182065, 40.45372203, 45.8866927 , 28.10829501, 42.97823461,
       40.17010842, 40.58478954, 45.90057583, 37.66793173, 34.47682269,
       29.09850847, 47.75208934, 53.405366  , 48.90580279, 40.18201292,
       37.77578944, 45.41366765, 37.31203242, 37.63105343, 45.3382713 ,
       44.57630487, 49.5728858 , 44.15311473, 31.94340407, 46.68592767,
       33.61122729, 41.14243218, 41.41473692, 43.99622056, 36.86653383,
       41.031416  , 45.22028501, 38.83016602, 45.39528278, 39.90373205,
       47.81545408, 43.97022153, 31.55574076, 40.084848  , 37.07212797]),
array([57.45861681, 56.7294408 , 55.31838943, 57.01690806, 57.21272736,
       69.13317575, 47.37561234, 58.54061452, 62.26952504, 54.23742909,
       54.9023783 , 48.79703193, 54.81616673, 55.33318052, 59.22467277,
       57.66540968, 51.52658061, 59.70251432, 68.10553365, 59.69868235,
       64.12760943, 58.05239207, 56.55060391, 57.5183316 , 59.52231118,
       58.85999019, 57.72726329, 53.17192917, 56.98594472, 52.54023738,
       48.7434834 , 61.78287849, 57.38210864, 67.53934455, 59.23418177,
       62.59123056, 55.24858148, 60.19012853, 58.65101959, 58.46370195]),
array([82.09095264, 76.94674302, 76.23136815, 84.68461279, 81.98877941,
       77.39800427, 74.7892971 , 78.1058941 , 78.41614261, 77.09975191,
       77.11473677, 79.4954805 , 75.18119658, 72.4505058 , 82.44328606,
       90.1469133 , 71.53353033, 77.70212125, 77.39195317, 86.30765655,
       74.75112345, 81.41918647, 74.42034576, 86.0447822 , 83.95341841,
       77.86457437, 88.08772398, 77.55388903, 80.46887916, 81.96182738,
       74.29429208, 83.30638025, 84.5160597 , 86.74719727, 91.89796225,
       85.69681906, 77.22915795, 79.96469162, 77.32218739, 79.46855256]),
array([102.51289854,  99.35400615,  98.27523846,  99.58519482,
       104.73923592,  99.98220718, 101.68204544, 100.40366474,
       107.93358149,  98.0170859 , 101.57546406,  93.10387687,
       109.20503798,  97.00227884,  95.55885471,  97.87732753,
        96.65369887, 101.09369775, 101.48337484,  95.80711622,
        97.00684364,  98.67787667,  97.22782552,  96.53693235,
        94.64583498,  98.47807505,  97.5229839 ,  98.23161063,
        91.13096547,  97.55223251,  96.18847461,  88.19459118,
        98.82867052, 100.02806462,  99.0436708 , 102.63665808,
```

```
       99.6279807 , 106.5554241 , 102.60782121, 102.01453367])]
********************************************************************************
**************************************
Cluster Sample [82.09095264 76.94674302 76.23136815 84.68461279 81.98877941
77.39800427
 74.7892971  78.1058941  78.41614261 77.09975191 77.11473677 79.4954805
 75.18119658 72.4505058  82.44328606 90.1469133  71.53353033 77.70212125
 77.39195317 86.30765655 74.75112345 81.41918647 74.42034576 86.0447822
 83.95341841 77.86457437 88.08772398 77.55388903 80.46887916 81.96182738
 74.29429208 83.30638025 84.5160597  86.74719727 91.89796225 85.69681906
 77.22915795 79.96469162 77.32218739 79.46855256 47.18182065 40.45372203
 45.8866927  28.10829501 42.97823461 40.17010842 40.58478954 45.90057583
 37.66793173 34.47682269 29.09850847 47.75208934 53.405366   48.90580279
 40.18201292 37.77578944 45.41366765 37.31203242 37.63105343 45.3382713
 44.57630487 49.5728858  44.15311473 31.94340407 46.68592767 33.61122729
 41.14243218 41.41473692 43.99622056 36.86653383 41.031416   45.22028501
 38.83016602 45.39528278 39.90373205 47.81545408 43.97022153 31.55574076
 40.084848   37.07212797]
```

Q4. Use the dataset given:

a) Create a random sample

[3]:
```
import pandas as pd

df1 = pd.read_excel('Dataset_01.xlsx')
print(df1.head())
```

```
   Distric_Sector ID - A single variable is created by combining the District ID
and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
0                                                111
1                                                111
2                                                111
3                                                111
4                                                111

   Grama Niladhari Division ID  Birth weight Mother's education  \
0                            1          3150          Secondary
1                            1          4180          Secondary
2                            1          2875          Secondary
3                            2          2840            Primary
4                            2          2540          Secondary

  Child's gender Wealth index quintile        Ethnicity Province Sector  \
0           Male                 Lowest  Sri Lanka Tamil  Western  Urban
1           Male                 Fourth          Sinhala  Western  Urban
2           Male                 Fourth  Sri Lanka Tamil  Western  Urban
3         Female                 Middle  Sri Lanka Tamil  Western  Urban
4           Male                 Fourth  Sri Lanka Tamil  Western  Urban
```

```
   Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
0                                                  0
1                                                  0
2                                                  0
3                                                  0
4                                                  0

   Mother's age Mother_Received_Thriposha  No. of clinical visits  \
0            33                      Yes                       9
1            30                      Yes                       6
2            32                      Yes                       7
3            29                      Yes                      10
4            37                      Yes                       6

   No. of months pregnant Mother's height     Mother's BMI
0                       9         Average  BMI 30 and over
1                       9         Average    BMI 18.5-24.9
2                       9         Average    BMI 18.5-24.9
3                       9           Short    BMI 24.9-29.9
4                       9            Tall    BMI 24.9-29.9
```

[4]:
```
random_sample = df1.sample(n=10, random_state=40)
random_sample
```

[4]:
```
      Distric_Sector ID - A single variable is created by combining the District
ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
1723                                                232
2318                                                332
2607                                                421
5123                                                912
341                                                 111
4423                                                712
375                                                 111
4352                                                712
1666                                                233
4647                                                812

      Grama Niladhari Division ID  Birth weight Mother's education  \
1723                         2365          2700             Higher
2318                         3329          3400          Secondary
2607                         4223          3000          Secondary
5123                         9071          2210          Secondary
341                           195          3105          Secondary
4423                         7179          2300          Secondary
375                           214          2600          Secondary
4352                         7151          3000          Secondary
```

```
1666                        2331          3000            Primary
4647                        8024          2600          Secondary

     Child's gender Wealth index quintile              Ethnicity  \
1723         Female                Second                Sinhala
2318           Male                Middle                Sinhala
2607           Male                Fourth  Sri Lanka moor /Muslim
5123         Female                Second                Sinhala
341            Male                Second                Sinhala
4423         Female                Fourth                Sinhala
375          Female                Fourth                Sinhala
4352         Female                Middle                Sinhala
1666           Male                Lowest           Indian Tamil
4647           Male                Lowest                Sinhala

           Province  Sector  \
1723        Central   Rural
2318       Southern   Rural
2607       Northern   Urban
5123   Sabaragamuwa   Rural
341         Western   Urban
4423  North-central   Rural
375         Western   Urban
4352  North-central   Rural
1666        Central  Estate
4647            Uva   Rural

      Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
1723                                                  0
2318                                                  0
2607                                                  0
5123                                                  1
341                                                   0
4423                                                  1
375                                                   0
4352                                                  0
1666                                                  0
4647                                                  0

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
1723            25                      Yes                      9
2318            32                      Yes                      6
2607            27                      Yes                      7
5123            39                      Yes                      5
341             30                      Yes                      6
4423            42                      Yes                     10
375             29                      Yes                      8
```

```
4352            38                      Yes                      8
1666            33                      Yes                      8
4647            32                      Yes                      7

      No. of months pregnant Mother's height     Mother's BMI
1723                      9            Tall       BMI < 18.5
2318                      9            Tall    BMI 24.9-29.9
2607                      9         Average       BMI < 18.5
5123                      9         Average    BMI 18.5-24.9
341                       9         Average  BMI 30 and over
4423                      9         Average    BMI 18.5-24.9
375                       9         Average    BMI 18.5-24.9
4352                      9         Average  BMI 30 and over
1666                      9           Short    BMI 18.5-24.9
4647                      9           Short  BMI 30 and over
```

b) Make a stratified sample based on BMI.

[5]:
```
stratum1 = df1[df1["Mother's BMI"]=="BMI 24.9-29.9"]
print("Strat1 with BMI 24.9-29.9")
display(stratum1)
print("*"*50)
random_strat1=stratum1.sample(n=3, random_state=30)
print("Random Sample of Strat1 with BMI 24.9-29.9")
display(random_strat1)
print("*"*70)

stratum2 = df1[df1["Mother's BMI"]=="BMI 18.5-24.9"]
print("Strat2 with BMI 18.5-24.9")
display(stratum2)
print("*"*50)
print("Random Sample of Strat2 with BMI 18.5-24.9")
random_strat2=stratum2.sample(n=3, random_state=30)
display(random_strat2)
print("*"*70)

stratum3 = df1[df1["Mother's BMI"]=="BMI 30 and over"]
print("Strat3 with BMI 30 and over")
display(stratum3)
print("*"*50)
print("Random Sample of Strat3 with BMI 30 and over")
random_strat3=stratum3.sample(n=3, random_state=30)
display(random_strat3)
print("*"*70)

stratum4 = df1[df1["Mother's BMI"]=="BMI < 18.5"]
print("Strat4 with BMI < 18.5")
```

```
display(stratum4)
print("*"*50)
print("Random Sample of Strat4 with BMI < 18.5")
random_strat4=stratum4.sample(n=3, random_state=30)
display(random_strat4)
print("*"*70)
```

```
Strat1 with BMI 24.9-29.9
```

```
      Distric_Sector ID - A single variable is created by combining the District␣
↪ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
3                                                    111
4                                                    111
11                                                   111
14                                                   111
15                                                   111
…                                                    …
5421                                                 922
5424                                                 922
5431                                                 922
5439                                                 923
5445                                                 922

      Grama Niladhari Division ID  Birth weight Mother's education  \
3                               2          2840            Primary
4                               2          2540          Secondary
11                              6          3600          Secondary
14                              6          3665          Secondary
15                              7          2340          Secondary
…                             …            …                  …
5421                         9288          3345          Secondary
5424                         9289          3100          Secondary
5431                         9292          2490          Secondary
5439                         9298          3000          Secondary
5445                         9300          2450          Secondary

     Child's gender Wealth index quintile              Ethnicity  \
3            Female                Middle         Sri Lanka Tamil
4              Male                Fourth         Sri Lanka Tamil
11           Female                Second  Sri Lanka moor /Muslim
14             Male                Middle  Sri Lanka moor /Muslim
15             Male                Middle                Sinhala
…               …                     …                     …
5421         Female                Middle           Indian Tamil
5424           Male                Middle                Sinhala
5431         Female                Lowest                Sinhala
5439           Male                Second                Sinhala
5445         Female                Lowest                Sinhala
```

```
          Province  Sector  \
3          Western   Urban
4          Western   Urban
11         Western   Urban
14         Western   Urban
15         Western   Urban
…            …     …
5421  Sabaragamuwa   Rural
5424  Sabaragamuwa   Rural
5431  Sabaragamuwa   Rural
5439  Sabaragamuwa  Estate
5445  Sabaragamuwa   Rural

      Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
3                                                     0
4                                                     0
11                                                    0
14                                                    0
15                                                    1
…                                                   …
5421                                                  0
5424                                                  0
5431                                                  1
5439                                                  0
5445                                                  1

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
3               29                      Yes                      10
4               37                      Yes                       6
11              37                      Yes                       7
14              26                      Yes                       8
15              28                      Yes                       6
…             …                       …                       …
5421            27                      Yes                      11
5424            33                      Yes                       7
5431            21                      Yes                       7
5439            28                      Yes                      11
5445            31                      Yes                      15

      No. of months pregnant Mother's height   Mother's BMI
3                          9           Short  BMI 24.9-29.9
4                          9            Tall  BMI 24.9-29.9
11                         9         Average  BMI 24.9-29.9
14                         9         Average  BMI 24.9-29.9
15                         9         Average  BMI 24.9-29.9
…                        …             …           …
5421                       9         Average  BMI 24.9-29.9
```

```
5424                        9         Average  BMI 24.9-29.9
5431                        9            Tall  BMI 24.9-29.9
5439                        9            Tall  BMI 24.9-29.9
5445                        9            Tall  BMI 24.9-29.9

[1572 rows x 16 columns]

************************************************
Random Sample of Strat1 with BMI 24.9-29.9

      Distric_Sector ID - A single variable is created by combining the District␣
 ↪ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
2003                                                 312
930                                                  132
1865                                                 312

      Grama Niladhari Division ID  Birth weight Mother's education  \
2003                         3091          2850            Primary
930                          1340          2615          Secondary
1865                         3030          2610          Secondary

     Child's gender Wealth index quintile              Ethnicity  Province  \
2003           Male                Second  Sri Lanka moor /Muslim  Southern
930            Male                Middle                 Sinhala   Western
1865         Female                Fourth                 Sinhala  Southern

     Sector  Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
2003  Rural                                                    0
930   Rural                                                    0
1865  Rural                                                    0

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
2003            36                      Yes                       7
930             27                      Yes                      12
1865            28                      Yes                       9

      No. of months pregnant Mother's height   Mother's BMI
2003                       9         Average  BMI 24.9-29.9
930                        9            Tall  BMI 24.9-29.9
1865                       9           Short  BMI 24.9-29.9

************************************************************************
Strat2 with BMI 18.5-24.9

      Distric_Sector ID - A single variable is created by combining the District␣
 ↪ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
1                                                    111
2                                                    111
8                                                    111
```

```
10                                                  111
19                                                  111
…                                                   …
5440                                                923
5441                                                923
5442                                                923
5443                                                923
5444                                                922

      Grama Niladhari Division ID  Birth weight Mother's education  \
1                               1          4180          Secondary
2                               1          2875          Secondary
8                               4          2670             Higher
10                              5          2655          Secondary
19                              8          3350             Higher
…                             …            …                 …
5440                         9298          2090            Primary
5441                         9298          2770          Secondary
5442                         9298          2850          Secondary
5443                         9298          2830          Secondary
5444                         9299          3700          Secondary

     Child's gender Wealth index quintile        Ethnicity      Province  \
1              Male                 Fourth          Sinhala       Western
2              Male                 Fourth  Sri Lanka Tamil       Western
8            Female                Highest  Sri Lanka Tamil       Western
10           Female                 Middle          Sinhala       Western
19           Female                Highest  Sri Lanka Tamil       Western
…             …                    …               …           …
5440         Female                 Lowest     Indian Tamil  Sabaragamuwa
5441         Female                 Second  Sri Lanka Tamil  Sabaragamuwa
5442           Male                 Lowest          Sinhala  Sabaragamuwa
5443         Female                 Lowest  Sri Lanka Tamil  Sabaragamuwa
5444         Female                 Lowest          Sinhala  Sabaragamuwa

      Sector  Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
1      Urban                                                          0
2      Urban                                                          0
8      Urban                                                          0
10     Urban                                                          0
19     Urban                                                          0
…      …                                                            …
5440  Estate                                                          1
5441  Estate                                                          0
5442  Estate                                                          0
5443  Estate                                                          0
5444   Rural                                                          0
```

```
      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
1               30                       Yes                       6
2               32                       Yes                       7
8               33                       Yes                       2
10              39                       Yes                       8
19              27                       Yes                       9
…              …                         …                        …
5440            36                       Yes                       9
5441            35                       Yes                       7
5442            29                       Yes                       7
5443            29                       Yes                      10
5444            34                       Yes                      15

      No. of months pregnant Mother's height   Mother's BMI
1                          9         Average  BMI 18.5-24.9
2                          9         Average  BMI 18.5-24.9
8                          9            Tall  BMI 18.5-24.9
10                         9         Average  BMI 18.5-24.9
19                         9         Average  BMI 18.5-24.9
…                         …              …            …
5440                       9           Short  BMI 18.5-24.9
5441                       9         Average  BMI 18.5-24.9
5442                       9           Short  BMI 18.5-24.9
5443                       9         Average  BMI 18.5-24.9
5444                       9         Average  BMI 18.5-24.9

[2651 rows x 16 columns]

**************************************************
Random Sample of Strat2 with BMI 18.5-24.9

      Distric_Sector ID - A single variable is created by combining the District␣
 ↪ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
164                                                   112
3542                                                  532
2180                                                  322

      Grama Niladhari Division ID  Birth weight Mother's education  \
164                            94          3400          Secondary
3542                         5333          3210          Secondary
2180                         3256          3625          Secondary

     Child's gender Wealth index quintile Ethnicity  Province Sector  \
164          Female                Fourth   Sinhala   Western  Rural
3542         Female                Middle   Sinhala   Eastern  Rural
2180         Female                Fourth   Sinhala  Southern  Rural

      Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
```

```
164                                                    0
3542                                                   0
2180                                                   0

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
164             37                      Yes                       8
3542            27                      Yes                       9
2180            27                      Yes                      10

      No. of months pregnant Mother's height   Mother's BMI
164                        9         Average  BMI 18.5-24.9
3542                       9         Average  BMI 18.5-24.9
2180                       9         Average  BMI 18.5-24.9
```

****

Strat3 with BMI 30 and over

```
      Distric_Sector ID - A single variable is created by combining the District␣
 ↪ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
0                                                    111
6                                                    111
7                                                    111
9                                                    111
12                                                   111
…                                                    …
5324                                                 922
5346                                                 922
5363                                                 922
5371                                                 922
5376                                                 922

      Grama Niladhari Division ID  Birth weight Mother's education  \
0                               1          3150          Secondary
6                               3          3000          Secondary
7                               3          2990          Secondary
9                               4          3300          Secondary
12                              6          2750          Secondary
…                             …            …                  …
5324                         9228          2550          Secondary
5346                         9243          2650          Secondary
5363                         9258          2200          Secondary
5371                         9262          4003          Secondary
5376                         9264          3560          Secondary

     Child's gender Wealth index quintile              Ethnicity  \
0              Male                 Lowest         Sri Lanka Tamil
6              Male                 Fourth  Sri Lanka moor /Muslim
7              Male                 Fourth  Sri Lanka moor /Muslim
```

```
9              Male               Highest         Sri Lanka Tamil
12           Female                Fourth  Sri Lanka moor /Muslim
…             …                    …                       …
5324           Male                Fourth                  Sinhala
5346           Male                Fourth                  Sinhala
5363           Male                Lowest                  Sinhala
5371           Male                Middle  Sri Lanka moor /Muslim
5376           Male                Second                  Sinhala

          Province Sector  \
0          Western  Urban
6          Western  Urban
7          Western  Urban
9          Western  Urban
12         Western  Urban
…            …    …
5324  Sabaragamuwa  Rural
5346  Sabaragamuwa  Rural
5363  Sabaragamuwa  Rural
5371  Sabaragamuwa  Rural
5376  Sabaragamuwa  Rural

      Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
0                                                     0
6                                                     0
7                                                     0
9                                                     0
12                                                    0
…                                                   …
5324                                                  0
5346                                                  0
5363                                                  1
5371                                                  0
5376                                                  0

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
0               33                      Yes                       9
6               41                      Yes                       8
7               24                      Yes                      10
9               39                      Yes                       7
12              24                      Yes                       5
…             …                       …                       …
5324            39                      Yes                       8
5346            31                      Yes                       6
5363            35                      Yes                       8
5371            38                      Yes                       7
5376            38                      Yes                       9
```

```
      No. of months pregnant Mother's height     Mother's BMI
0                          9         Average  BMI 30 and over
6                          9         Average  BMI 30 and over
7                          9            Tall  BMI 30 and over
9                          9            Tall  BMI 30 and over
12                         9            Tall  BMI 30 and over
…                        …              …               …
5324                       9            Tall  BMI 30 and over
5346                       9            Tall  BMI 30 and over
5363                       9         Average  BMI 30 and over
5371                       9         Average  BMI 30 and over
5376                       9            Tall  BMI 30 and over

[584 rows x 16 columns]

**********************************************
Random Sample of Strat3 with BMI 30 and over

      Distric_Sector ID - A single variable is created by combining the District␣
 ↪ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
2588                                                421
1755                                                231
1236                                                212

      Grama Niladhari Division ID  Birth weight Mother's education  \
2588                         4211          3720          Secondary
1755                         2383          2750          Secondary
1236                         2049          2370          Secondary

     Child's gender Wealth index quintile              Ethnicity  Province  \
2588         Female                Middle         Sri Lanka Tamil  Northern
1755         Female                Middle  Sri Lanka moor /Muslim   Central
1236           Male               Highest                 Sinhala   Central

     Sector  Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
2588  Urban                                                          0
1755  Urban                                                          0
1236  Rural                                                          1

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
2588            29                      Yes                       8
1755            41                       No                       7
1236            27                      Yes                       9

      No. of months pregnant Mother's height     Mother's BMI
2588                       9         Average  BMI 30 and over
1755                       9         Average  BMI 30 and over
1236                       9         Average  BMI 30 and over
```

```
**********************************************************************
Strat4 with BMI < 18.5

      Distric_Sector ID - A single variable is created by combining the District␣
 ↪ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
5                                                      111
31                                                     111
42                                                     111
73                                                     112
83                                                     112
…                                                      …
5407                                                   923
5409                                                   922
5412                                                   922
5433                                                   923
5437                                                   923

      Grama Niladhari Division ID  Birth weight Mother's education  \
5                               2          3070             Higher
31                             17          2800          Secondary
42                             22          2000          Secondary
73                             36          3270             Higher
83                             40          2130          Secondary
…                             …            …                …
5407                         9280          2170          Secondary
5409                         9281          2540          Secondary
5412                         9285           400          Secondary
5433                         9293          2900            Primary
5437                         9296          2400            Primary

     Child's gender Wealth index quintile              Ethnicity  \
5              Male                 Fourth  Sri Lanka moor /Muslim
31           Female                 Second  Sri Lanka moor /Muslim
42           Female                 Fourth  Sri Lanka moor /Muslim
73             Male                Highest                 Sinhala
83             Male                 Middle            Indian Tamil
…               …                   …                      …
5407         Female                 Lowest         Sri Lanka Tamil
5409         Female                 Second                 Sinhala
5412           Male                 Middle                 Sinhala
5433         Female                 Second         Sri Lanka Tamil
5437           Male                 Lowest         Sri Lanka Tamil

          Province  Sector  \
5          Western   Urban
31         Western   Urban
42         Western   Urban
73         Western   Rural
```

```
83         Western   Rural
…            …     …
5407  Sabaragamuwa  Estate
5409  Sabaragamuwa   Rural
5412  Sabaragamuwa   Rural
5433  Sabaragamuwa  Estate
5437  Sabaragamuwa  Estate

      Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
5                                                     0
31                                                    0
42                                                    1
73                                                    0
83                                                    1
…                                                   …
5407                                                  1
5409                                                  0
5412                                                  1
5433                                                  0
5437                                                  1

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
5               24                        No                       7
31              21                       Yes                       6
42              24                       Yes                       8
73              31                       Yes                      14
83              19                       Yes                      10
…              …                         …                      …
5407            18                       Yes                       3
5409            19                       Yes                       7
5412            30                       Yes                       5
5433            31                       Yes                       7
5437            22                       Yes                       6

      No. of months pregnant Mother's height Mother's BMI
5                          9         Average   BMI < 18.5
31                         9         Average   BMI < 18.5
42                         9         Average   BMI < 18.5
73                        10            Tall   BMI < 18.5
83                        10            Tall   BMI < 18.5
…                        …              …          …
5407                       9         Average   BMI < 18.5
5409                       9         Average   BMI < 18.5
5412                       9            Tall   BMI < 18.5
5433                       9         Average   BMI < 18.5
5437                       9         Average   BMI < 18.5

[639 rows x 16 columns]
```

```
************************************************
Random Sample of Strat4 with BMI < 18.5

      Distric_Sector ID - A single variable is created by combining the District␣
 ↪ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
1898                                                312
2943                                                452
3884                                                612

      Grama Niladhari Division ID  Birth weight Mother's education  \
1898                         3046          2750          Secondary
2943                         4516          3600          Secondary
3884                         6116          2760             Higher

     Child's gender Wealth index quintile        Ethnicity       Province  \
1898           Male                Middle          Sinhala       Southern
2943           Male                Lowest  Sri Lanka Tamil       Northern
3884         Female               Highest          Sinhala  North-western

     Sector  Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
1898  Rural                                                          0
2943  Rural                                                          0
3884  Rural                                                          0

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
1898           30                       Yes                       8
2943           24                       Yes                       6
3884           38                       Yes                       9

      No. of months pregnant Mother's height Mother's BMI
1898                       9         Average   BMI < 18.5
2943                       9         Average   BMI < 18.5
3884                       9            Tall   BMI < 18.5

************************************************************************
```

c) Make cluster sample based on province.

[8]:
```
import numpy as np

cluster1 = df1[df1["Province"]=="Western"]
cluster2 = df1[df1["Province"]=="Sabaragamuwa"]
cluster3 = df1[df1["Province"]=="Central"]
cluster4 = df1[df1["Province"]=="Eastern"]
cluster5 = df1[df1["Province"]=="North-central"]
cluster6 = df1[df1["Province"]=="North-western"]
cluster7 = df1[df1["Province"]=="Uva"]
cluster8 = df1[df1["Province"]=="Northern"]
cluster9 = df1[df1["Province"]=="Southern"]
```

```
cluster=[cluster1,cluster2,cluster3,cluster4,cluster5,cluster6,cluster7,cluster8,cluster9]
chosen_index1=np.random.choice(len(cluster),size=2,replace=False)
print(chosen_index1)

cluster_sample2=pd.concat([cluster[i] for i in chosen_index1])
cluster_sample2
```

```
[0 3]
```

[8]:

	Distric_Sector ID - A single variable is created by combining the District ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)	Grama Niladhari Division ID	Birth weight	Mother's education	Child's gender	Wealth index quintile	Ethnicity	Province	Sector
0	111	1	3150	Secondary	Male	Lowest	Sri Lanka Tamil	Western	Urban
1	111	1	4180	Secondary	Male	Fourth	Sinhala	Western	Urban
2	111	1	2875	Secondary	Male	Fourth	Sri Lanka Tamil	Western	Urban
3	111	2	2840	Primary	Female	Middle	Sri Lanka Tamil	Western	Urban
4	111	2	2540	Secondary	Male	Fourth	Sri Lanka Tamil	Western	Urban
…	…	…	…	…	…	…	…	…	…
3612	532	5359	3140	Secondary	Female	Lowest	Sri Lanka Tamil	Eastern	Rural
3613	532	5360	1150	Primary	Female	Lowest	Sri Lanka Tamil	Eastern	Rural
3614	532	5360	3460	Secondary	Male	Lowest	Sri Lanka Tamil	Eastern	Rural
3615	532	5360	2600	Primary					
3616	532	5360	2910	Secondary					

```
3615         Female                Second  Sri Lanka Tamil  Eastern  Rural
3616           Male                Lowest  Sri Lanka Tamil  Eastern  Rural

      Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
0                                                     0
1                                                     0
2                                                     0
3                                                     0
4                                                     0
…                                                   …
3612                                                  0
3613                                                  1
3614                                                  0
3615                                                  0
3616                                                  0

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
0               33                      Yes                       9
1               30                      Yes                       6
2               32                      Yes                       7
3               29                      Yes                      10
4               37                      Yes                       6
…              …                        …                       …
3612            26                      Yes                       7
3613            23                      Yes                       3
3614            26                      Yes                       8
3615            30                      Yes                       6
3616            28                      Yes                       8

      No. of months pregnant Mother's height     Mother's BMI
0                          9         Average  BMI 30 and over
1                          9         Average    BMI 18.5-24.9
2                          9         Average    BMI 18.5-24.9
3                          9           Short    BMI 24.9-29.9
4                          9            Tall    BMI 24.9-29.9
…                        …              …              …
3612                       9            Tall    BMI 24.9-29.9
3613                       7         Average    BMI 18.5-24.9
3614                       9         Average    BMI 18.5-24.9
3615                       9         Average    BMI 18.5-24.9
3616                       9         Average    BMI 18.5-24.9

[1722 rows x 16 columns]
```

d) Systematic sample based on mothers age (choose interval on your preference)

[9]:
```
interval=50
sys_sample = df1.iloc[::interval]

# Display the sample
display(sys_sample)
```

```
     Distric_Sector ID - A single variable is created by combining the District␣
↪ID and the Sector ID (Urban - 1, Rural - 2, Estate - 3)  \
0                                                    111
50                                                   111
100                                                  111
150                                                  112
200                                                  112
…                                                    …
5200                                                 912
5250                                                 912
5300                                                 922
5350                                                 922
5400                                                 923

      Grama Niladhari Division ID  Birth weight Mother's education  \
0                               1          3150          Secondary
50                             26          2320          Secondary
100                            54          3720             Higher
150                            89          2290          Secondary
200                           113          2670          Secondary
…                             …            …             …
5200                         9099          2220          Secondary
5250                         9118          2960          Secondary
5300                         9215          3500          Secondary
5350                         9246          3110          Secondary
5400                         9276          2750          Secondary

     Child's gender Wealth index quintile              Ethnicity  \
0              Male                 Lowest        Sri Lanka Tamil
50           Female                Highest  Sri Lanka moor /Muslim
100          Female                 Fourth                Sinhala
150          Female                 Middle                Sinhala
200            Male                Highest                Sinhala
…               …                  …                     …
5200         Female                 Second                Sinhala
5250         Female                 Middle                Sinhala
5300           Male                Highest                Sinhala
5350           Male                 Fourth                Sinhala
5400         Female                 Lowest        Sri Lanka Tamil

          Province  Sector  \
```

```
0          Western   Urban
50         Western   Urban
100        Western   Urban
150        Western   Rural
200        Western   Rural
…            …     …
5200  Sabaragamuwa   Rural
5250  Sabaragamuwa   Rural
5300  Sabaragamuwa   Rural
5350  Sabaragamuwa   Rural
5400  Sabaragamuwa  Estate

      Low birth weight - 1 if it is a low birth weight, 0 otherwise  \
0                                                          0
50                                                         1
100                                                        0
150                                                        1
200                                                        0
…                                                        …
5200                                                       1
5250                                                       0
5300                                                       0
5350                                                       0
5400                                                       0

      Mother's age Mother_Received_Thriposha  No. of clinical visits  \
0               33                      Yes                       9
50              28                      Yes                      10
100             36                      Yes                      10
150             27                      Yes                       9
200             31                      Yes                       9
…             …                       …                       …
5200            30                      Yes                       8
5250            31                      Yes                       7
5300            43                      Yes                       9
5350            30                      Yes                       8
5400            25                      Yes                       8

      No. of months pregnant Mother's height     Mother's BMI
0                          9         Average  BMI 30 and over
50                         9            Tall    BMI 24.9-29.9
100                       10         Average    BMI 24.9-29.9
150                        9           Short    BMI 18.5-24.9
200                        9         Average    BMI 18.5-24.9
…                        …             …              …
5200                       9         Average       BMI < 18.5
5250                       9         Average    BMI 18.5-24.9
5300                       9            Tall    BMI 24.9-29.9
```

```
5350                      9         Average    BMI 18.5-24.9
5400                      9         Average       BMI < 18.5

[109 rows x 16 columns]
```