Data set: [Popular Baby Names (kaggle.com)](https://www.kaggle.com/datasets/utkarshx27/popular-baby-names?resource=download)

Department of Health and Mental Hygiene (DOHMH) is the data source.

State of the Union (US)



I don’t know how to connect them together to make a story telling report, I tried my best to somewhat link them.

Chapter 1:

We have a total of 606,104 individuals in the sample.

And we have a total of 18,053 different combinations.

With those, we can say that the mean is , which means that we have an average of 34 names per combination.

We would then find the variance of the sample, using

Which results in 1496. And the SD would be

----------------------------------------------------------------------------------------------------------

2.4. A couple who just got a baby girl and is picking her a name, they always love Hispanic names and deciding to pick it. Given the list of names to choose from. What’s the probability at random that the couple would pick the name Grace?

From data set, there a total of 83,589 individuals with female Hispanic names from year 2011 to 2019.

Since the sample point is vast, I cannot list all the events in the sample point. I can only list the events where Grace was chosen. And it’s probability.

The events for each year respectively where “Grace” was chosen are:

E1 = , E2 = , E3 = , E4 = , E5 = , E6 = , E7 = , E8 = , E9 =

So, the probability of the couple randomly picks the name “Grace” is

----------------------------------------------------------------------------------------------------------

A couple is expecting a twin, boy and girl. Based on the data set, how many possible ways can the couple choose?  
the number of boy names: 8874

the number of girl names: 9179

using the m x n rule.

there are 8874 x 9179 = 81, 454, 446 possible combinations of boy and girl name

----------------------------------------------------------------------------------------------------------

Another couple with another twin, 2 boys this time and they want to pick 2 names out of the 8874 names. The couple narrowed it down to 15 names that they want to pick. Now, they want to pick those 15 names randomly from a box, without replacement. Whatever the first name they picked, will be the older brother. How many possible ways can the names be picked?

Using permutation:

We calculate the number of sample points

ways of picking names

----------------------------------------------------------------------------------------------------- -

Now, the wife suddenly felt a kick in her stomach, she realized that her babies heard what they were saying and disagreed with the way they decide who is older. So now the couple has decided that no matter what name they picked first, it won’t have any decision on who is older, and let it be decided in the battleground that is the womb. How many possible ways can they pick 2 names now?

In this instance, we would use combination since order don’t matter now:

possible ways of picking 2 names.

From the data set, narrowed to the year 2013,

Given that the probability of the name “Eli” is chosen: P(A):

And the probability that the chosen name is in the year: P(B): 2013 :

And the probability that the name “Eli” is chosen in the year 2013: P(A B):

Find:

1. P(A|B)
2. P(B|A)
3. P(A|A U B)
4. P(A|B) =
5. P(B|A) = 0.96
6. P (A|A U B) =

P(A U B) =

Suppose a couple wanting to pick a name for their new born baby girl, and they really don’t want to pick “Gia” or “Haylee”. Whats the probability of of picking a name that is not “Gia” or “Haylee” at random?

P(Gia) =

P(Haylee) =

P(A) = 1 – (0.0012 + 0.0009) = 0.998 or 99.8% that they would pick a name other than Gia or Haylee.

A clinic is conducting a survey on the data of babies born at their facility. This part of the survey specifically focuses on babies born in 2011 who are Hispanic. The clinic wants to determine the probability that a randomly selected baby's name starts with the letter "G".

We would use the theorem of total probability here:

P(A) = the probability that the baby was born in 2011 and is Hispanic

P(B) = the probability of a random baby whose name start with the letter “G”

P(A|B) = the probability that a baby was born in 2011 and is Hispanic given that the baby’s name starts with the letter “G”

From the data set we know that

Bayes theorem to find P(B|A) and P(B|

P(B) = 0.01932

A couple is having a twin, and is open with name picking, they want to pick 3 names out of a range of 4 names that they like, 2 are male names and 2 are female names. They decided to name their babies regardless of name genders. Finally, they put all the names in a box and pick randomly. Let Y be the number of picking male name. What’s the probability distribution for Y?

We would want to use the probability distribution formula:

Y=0, 1, 2

The expected value for this is

Variance is

A couple is planning to have a family gathering to celebrate their upcoming family addition. During the gathering, the family play a game to decide the name of the baby. A game of tossing 5 coins into the air. The number of head will decide the father or the mother be the one to pick the name. If the number of head is 3 or more, the mother pick the name. Find the probability that the father would be the one to pick the name. Knowing that the father prepare the coins, we know that the chances of getting tail is 0.6.

We would use binomial distribution to solve this problem:

p = 0.6, q = 0.4

y = 3, 4, 5 tails

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| y | 3 | 4 | 5 | total |
|  |  |  |  |  |
|  | 0.345 | 0.259 | 0.0777 | 0.6817 |

The expected is E(Y) = np = 5\*0.6 = 3

Variance: V(Y) = npq = 5 \* 0.6 \* 0.4 = 102

Moreover, the probability of the first tail at Y = 1, 2, 3, 4, 5 attempt are:

p(y) = qy-1p

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Y | 1 | 2 | 3 | 4 | 5 |
|  | 0.6 | 0.24 | 0.096 | 0.0384 | 0.01536 |

Expected:

Variance:

While we at it, the father also want to know the probability of the successes (r = 3) in y = 3, 4, 5 tosses. To further his belief that he’s going to be the one to name the baby.

We would use negative binomial to find the 3rd success in y trials

r = 3

|  |  |  |  |
| --- | --- | --- | --- |
| y | 3 | 4 | 5 |
|  |  |  |  |
|  | 0.216 | 0.259 | 0.415 |

Expected:

Variance:

A Hospital is collecting data from the same data set, from a randomly selected 70 baby names, 10 from each ethnicity. They then pick at random 20 names. What’s the probability that they would pick all 10 Hispanic names in the 20 selected?

Use would use hypergeometric distribution for this one:

Expected:

Variance

From the data set, the Department of Health and Mental Hygiene get on average 8 baby’s names per hour in 2014. What’s the probability that they obtain 12 baby names per hour.

We want to use Poisson distribution for this one:

K = 12

According to the data set, in the year 2014, we have the average of 34 baby that have that name. And the standard deviation is 4. Lets estimate the proportion of baby names that whose count lies within 3 standard deviations from the mean.

We would use Tchebyshev’s theorem:

Mean = 34

SD = 4

K = 3

I don’t know how to apply my data set to probability distribution and density function for a continuous random variable. I mean the numbers in my data set aren’t random.

I can only mimic examples in the book for this one.

(4.2)

A gacha vending machine is located in the middle of a mall, the box has 5 type of items in it. In there is a figurine you want. The gacha randomly selects an item (no repetition) when you win the gacha game. You play the game and get the items one at a time until you get your wanted figurine. Let Y be the number of trials on which you won the item you want.

1. Find the probability function for Y?
2. Give the corresponding distribution function.
3. What is P(Y < 3), P(Y ≤ 3), P(Y = 3)?

, you looking for the first success within 5 trials, thus

Same for this one, I don’t know how to apply it. Or any integration into my data set tbh.

(4.14)

Find F(y)

Graph both

Expected:

Variance:

Lets say that 10 names are selected, the number of babies with one of those names can range from 0 to 135.

We want to find.

1. The probability that the count is under 34
2. Above 78
3. The expected and variance

Lets find them using uniform distribution means

a = 0, b = 135

a.

b.

c.

Expected:

Variance:

The clinic records the babies’ name using the data set and found that it have an exponential distribution with mean of 1.6. Find the probability that the data set changes/added thus make it:

1. Exceed 2.5 measurement
2. Fall between 1.2 and 2
3. Find expected and variance

By using exponential distribution,

a)

b)

c)

expected: 1.6

Variance:1.6­­­­2 = 2.56

From the data set, 10 names are pulled and put in a box, 2 names are selected randomly and drop in either box A, box B or the trash.

Let x = the number of names in box A

Y = the number of names in box B

1. Find the joint probability of x and y:
2. Find F(1,0)

We use Multivariate probability function here:

a.

Lets list all events of x and y:

S = {(0,0), (1,0), (2,0), (1,1), (0,1), (0,2)}

Chances are

B)

I don’t know how to come up with a joint density function so I’m just going to use one of the book’s problems.

(5.26)

1. Find the marginal density function of x and y
2. The conditional density function of x given Y = y
3. The conditional density function of y given X = x

a.

To find the marginal density function, we use:

According to the independent law, we find that the marginal density function of x and y and the joint density function to be independent because:

b.

find

numerator:

Denominator:

c.

The conditional density function of x given Y = y

d.

The conditional density function of x given Y = y