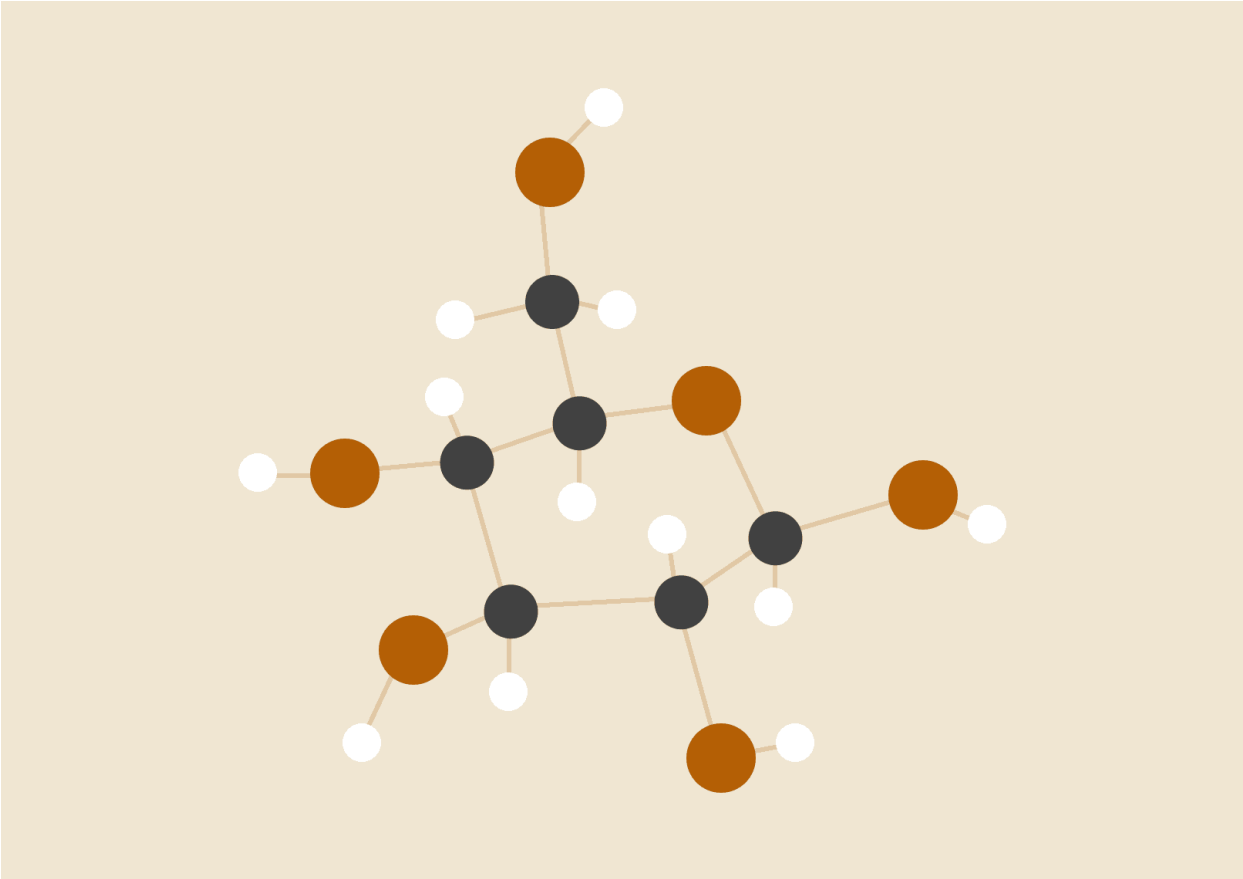
Baby Names and Numbers



Jack Stewart

12.10.2024
Probability and Applied Statistics

Problems in this paper are based on data from the csv file found in the Part2 folder, which is a csv file containing a list of baby's Names, Gender, Ethnicity, Count, and Rank for a specific Year, for the years 2011 through 2016.

2.3: A review of Set Notation

Suppose a family picks two baby names. With F representing a Girl name and M representing a boy name, and a pair {XY} letting X represent the first picked and Y being the second picked, such that the set of possibilities is S = {FF, FM, MF, MM}.

Let A mean there are no Females, B means there is at least one male, and C means both children are different genders.

What are the probabilities of A, B, C, AnB, anC, BuC?

P(A) = ¼, because it is the MM Case

P(B) = ¾, because it is true for all cases besides FF

P(C) = ½, because in half of the cases it is true (MF and FM)

P(AnB) = ¼, because this is true for the case of MM

p(AnC) = 0, because this is not true for any case

p(BnC) = ½, because this is true for MF and FM

2.4: Experiments with Discrete Cases

Every childname in the dataset is assigned an ethnicity of either Hispanic, White, Asian and Pacific Islander, or Black. Let these equal H, W, A, and B respectively. Each name is also Male or Female, let these be represented by M, F respectively. List the sample space for the experiment of choosing one baby name with some given ethnicity and gender.

SS = {HM WM AM BM HF WF AF BF}

2.6: Tools for Counting Sample Points

A concerned parent cannot decide between 4 different baby names for their son. They ask their brother for input, and he lists the 4 names in order from his favorite to his least favorite. How many different orderings could he have made?

Size of SS = 4! = 4*3*2*1 = 24

2.7: Conditional Probability and Independence

The probability A baby's name is hispanic is 25% and the probability that a baby's name is hispanic and starts with the letter G is 1.375%. What is the probability of a baby's name starting with G given that baby is hispanic?

The probability is 1.375% / 25% = 5.5%

2.8: Two Laws of Probability

A Baby Name is going to be selected randomly. The probability of one baby name being hispanic is 25% and the probability of a baby name starting with the letter G given it is hispanic is 5.5%. The overall probability of selecting a hispanic baby name starting with G is 1.375%. Are the events of selecting a hispanic baby name and that name starting with G independent?

Yes. 25% * 5.5% = 1.375% so the events are independent.

2.10: Bayes Rule

Suppose that of the baby names starting with K, 60% are male and 40% are female. Of the male babies whose names start with K, 1% are named Kayn but of the female babies only 0.5% are named Kayn. What are the odds that a baby is male given it is named Kayn?

P(M) = there is a 75% chance the baby is male, 75% = 60% * 1% / (60% * 1% + 40% * .5%)

3.3: Expected Value and Average

The number of times the name Jeremy was seen among Hispanics every year is recorded

below, Find the expected value of the number, if a year was selected at random.

2011	2012	2013	2014	2015	2016
4	6	4	5	7	7

Expected Value = (4 + 6 + 4 + 5 + 7 + 7) / 6 = 5.5

3.4: Binomial Distribution

10 Hispanic Babies are selected from the US randomly. Among hispanic babies there is a 5.5% chance that their name starts with a G. What are the odds that at least 2 of the 10 have names starting with G?

P(y >= 2) = P(y=2) = P(y = 3) . . . P(y = 10)

P(y = x) = 5.5%^x + (1-5.5%)^(10 - x)(10 Choose 5)

p(y >= 2) = 0.10148 = 10.148%

3.5: Geometric Distribution

Hispanic babies are drawn out of a hat which has a sufficient supply of babies until one of them has a name starting with G, which is a 5.5% chance. What are the odds one of the first 10 hispanic babies drawn will have a name starting with G?

P(y >= 10) = 1 - (1-5.5%)^10 = 43%

3.7: HyperGeometric Distribution

10 Babies are selected to wait to be selected randomly. 70% of these babies are male and 30% are female. 3 Babies are selected, what are the odds that they are all male?

P(All Male) = 7/10 * 6/9 * 5/8 = 0.292

3.8: Poisson Distribution

Suppose that there are on average 2 babies named John born in Delaware each second. What is the probability that there will be 4 Babes named John born in a specific second?

P(y = 4) = 2^4/4! * e^-(2) = 9.022%

4.2: Density Function vs Distribution Function

Suppose the probability of a baby being White based on their birthplace was represented by x, where x is the distance to wyoming in miles. What is the probability density function of being white based on distance to wyoming?

F(y) = x^2, f(y) = F'(x) = 2x

4.3: Expected Values of Continuous Random Variables

If the distribution of a baby's name is 2x based on the parent's age, find the Expected value, 20 <= x <= 50

F(x) = 2x, f(x) = 2. The integral of 2y dy from 20 to 50 is 50^2 - 20^2 which is 2100.