Probability and Statistics of Pokémon

*Database source link:* [*https://www.kaggle.com/datasets/rounakbanik/pokemon*](https://www.kaggle.com/datasets/rounakbanik/pokemon)

This project was an interesting way to investigate database research. I honestly did not expect it to be so easy to find open databases, nor did I expect there to be a thriving community around them such as those seen on Kaggle. In using Kaggle, I was able to discover plenty of online databases to fit my interests and decided to settle on one focused on Pokémon. The Pokémon games have a lot of statistical research poured into them by adoring fans. The statistics within Pokémon can be understood and manipulated to create the perfect team for regular gameplay or competitive battling. Sometimes, these statistics can be manipulated to provide a higher chance of discovering Shiny Pokémon, which is very sought after by any Pokémon fan. I picked this database because it seemed to have the most comprehensive data regarding each individual Pokémon and did not rely on specific formulas used in any of the games. Thus, this database provided the most information and flexibility to mold questions around.

In creating questions surrounding this Pokémon database, I realized how challenging it can be to create mathematical problems. At times, the earlier sections were easier to form questions around than the later sections. This could be because the earlier sections were somewhat easier than later sections, but this is also a result of the later sections becoming more and more specific to certain problems. While this database did provide some flexibility, there were some problems that I felt were not easy to form around my given data, and thus had to be restricted further. I tried to provide as much context as possible for each problem to make sense to a layman. By the end of this process, I’ve discovered how easy it can be to game the shiny hunting and hatching mechanics of Pokémon eggs. In doing additional background research, I’ve also learned a lot more about other randomized mechanics in the game, but their formulas are often very complex and change depending on the game. In this regard, I was pleasantly surprised to know that the developers of Pokémon games, GameFreak, put in extra effort to fine tune these randomized events and their statistical outputs.

1. [Set Notation] Your team of Pokémon have the types of Poison, Dark, Flying, Grass, Water, and Normal. Your brother’s team of Pokémon have the types of Psychic, Flying, Electric, Fire, Dragon, and Dark. Given the total available types are S = {Normal, Fire, Grass, Water, Electric, Ice, Fighting, Poison, Ground, Flying, Psychic, Bug, Rock, Ghost, Dark, Dragon, Steel, Fairy}…
   1. Q: What are the remaining available types?  
      A: Ice, Fighting, Ground, Bug, Rock, Ghost, Steel, Fairy
   2. Q: Which types are shared between you and your brother’s teams?  
      A: Flying, Dark
2. [Mean/Median] You have a team of Pokémon consisting of Watchog, Stoutland, Audino, Cinccino, and Bouffalant. These Pokémon weigh 27.0 kg, 61.0 kg, 31.0 kg, 7.5 kg, and 94.6 kg, respectively.
   1. Q: What is the average weight of your team?  
      A: 27.0 + 61.0 + 31.0 + 7.5 + 94.6 = 221.1 / 5 = 44.22 kg
   2. Q: Which Pokémon is the median of the team by weight?  
      A: In order by ascending weight, the team is: Cinccino (7.5 kg), Watchog (27.0 kg), Audino (31.0 kg), Stoutland (61.0 kg), and Bouffalant (94.6 kg). So, Audino is the median.
3. [Probability Introduction] You breed 3 Pokémon with a 50% chance of being female and 50% chance of being male.
   1. Q: What are the odds of hatching exactly 2 females?   
      A: Given these conditions, there will be 8 possible outcomes as follows:  
      E1: FFF E2: FFM E3: FMF E4: MFF  
      E5: FMM E6: MFM E7: MMF E8: MMM  
      Each outcome is equally likely, so they all have a probability of 1/8. E2, E3, and E4 each have exactly 2 females, so P(A) = P(E2) + P(E3) + P(E4) = 1/8 + 1/8 + 1/8 = 3/8
4. [Multinomial Coefficients] You and three of your friends begin your new Pokémon journey. You begin your journey by selecting a team of 6 Pokémon from the starting area.
   1. Q: Given 24 available Pokémon, how many distinct team configurations of 6 Pokémon can be created?  
      A:
5. [Conditional Probability] You decide to breed a Vulpix and an Eevee for your team. Vulpix has a 25.6% chance of hatching as male, while Eevee has a 88.1% chance of hatching as male. Assuming the probability of both Vulpix and Eevee hatching as male is 40%...
   1. Q: What is the probability that Eevee hatches male given that Vulpix hatches male?  
      A: Given P(Male Vulpix) = 0.256, P(Male Eevee) = 0.881, and P(Male Vulpix ∩ Male Eevee) = 0.40, we can set up the conditional probability equation as follows:
6. [Independent Events] When encountering wild Pokémon, there is a chance that Pokémon will be a shiny, meaning it will appear with different colors. Assuming that encounters are independent events…
   1. Q: What is the probability of encountering a shiny Pokémon in 5 encounters?  
      A: Since these events are independent, we can multiply probabilities together. So,
7. [Theorem of Total Probability & Multiplicative Law of Probability] To discover new Pokémon, you must encounter them in wild areas such as routes between towns. Route 1 has 4 Normal type Pokémon, and 4 non-Normal type Pokémon. Route 2 has 1 Normal type Pokémon, and 10 non-Normal type Pokémon. Route 3 has 7 Normal type Pokémon, and 11 non-Normal Pokémon. You decide to randomly pick one of these routes to travel down and begin a random encounter there. Assume all routes and Pokémon type encounters are equally likely to occur.
   1. Q: What is the probability that the Pokémon encountered is a Normal type?  
      A: By the Theorem of Total Probability, .   
      So we have .   
      Given the Multiplicative Law of Probability, we can expand this to be
8. [Bayes Theorem] Referencing Question 7…
   1. Q: The Pokémon encountered is a Normal type. What is the probability you encountered it on Route 2?  
      A: By Bayes Theorem,   
      So,
9. [Binomial Distribution] Recall that when encountering wild Pokémon, there is a chance that Pokémon will be shiny. You decide to hunt for shiny Pokémon in the Safari Zone, a special area that limits you to only 50 encounters per visit.
   1. Q: What is the probability of encountering at least one shiny?  
      A: We can use the Binomial Distribution where n = 50, p = , and q = , to solve for when y = at least one shiny, or 1 – p(0).   
      Thus,   
      So, you have a 0.61% chance of encountering a shiny during your visit to the Safari Zone.
10. [Geometric Distribution] Referencing Question 9…
    1. Q: What is the probability that you find a shiny Pokémon on your 25th encounter in the Safari Zone?  
       A: Using the Geometric Distribution, we still have that p = , and q = , but we now have that n = 25.   
       So, =   
       Meaning you have a 0.01% chance to encounter a shiny in your 25th encounter.
11. [Hypergeometric Distribution] You decide to make a trip to the Safari Zone again, limiting yourself to 50 random counters. Before entering, you are told that there are 30 Grass type Pokémon and 20 Water type Pokémon inside. During this trip, you end up catching 10 Pokémon.
    1. Q: What is the probability that exactly 5 of these captured Pokémon are Water type?  
       A: In this example, we can see that N = 50, n = 10, r = 20, y = 5, (N-r) = 30, and (n-y) = 5. Using the Hypergeometric Distribution, we can see that or 0.0492% chance that exactly 5 of the captured Pokémon are Water types.
12. [Tchebysheff’s Theorem] When encountering a Pokémon in any given area, the Pokémon’s level can range depending on the location. In the Wild Area, a Pokémon’s level can range from level 5 to level 55. Given that the mean level is 30 and the standard deviation is 4,
    1. Q: What percentage of Pokémon will spawn with a level between 20 and 40?  
       A: Given that 40 – 30 = 10 and 30 -   
       So, or 84% of Pokémon between level 20 and 40.
13. [Poisson Distribution] When traveling through an area in a 5-minute period, you have an 80% chance of encountering random Pokémon by the Poisson Distribution.
    1. Q: If you travel through an area for 5-minutes, what is the probability that you will encounter 3 Pokémon?  
       A: Given that λ = 0.80 and y = 3, then by the Poisson Distribution we have: or 0.4248% of encountering 3 Pokémon.