An Comparison of Statistical Analysis In Minecraft Speedrunning

**STAT 3910**  
**Probability and Mathematical Statistics II (Fall 2023)**  
**Final Project**  
**Name: Hashem Fawzy**

**Problem Setting**: Minecraft is considered one of the most popular video game ever made. One aspect of the game that is completed regularly every day by players is called speed-running (I.E, the act of completing the game as fast as possible). To complete the game, the fastest method completed on a version of the game (1.16) includes the act of trading gold items to a game mob called a Piglin for a chance to get Ender Pearls, kill a hostile mob called a Blaze for a chance to get items called blaze rods, then use those two items to find the final boss of the game called the Ender Dragon, beat it, then win the game.

In 2020, a speed-run of a game on version 1.16 of Minecraft had been conducted by a player named Clay, AKA Dream. he had conducted the acts mentioned above, completed the run, and submitted it to the official speed running forums. 2021 had a player find that during that speed run, the two acts, finding Ender Pearls, and collecting Blaze Rods during the run submitted had found the odds to be particularly lucky, almost too lucky. Comparing 6 other consecutive runs, they had found his odds during those runs of the two events to be as big as Billion of receiving luck such as that. Those odds are considered too high to even be possible, as a result, the moderators had decided that Dream might have cheated to increase his odds. This was a common occurance in Minecraft speedrunning, as not only is the efforts to maintain a fast and efficient run plagued by random possibilities ruining any chance of a personal best time, but techniques to cheat were more easier to conduct for many and harder to detect as more and more speedrunning times were submitted.

His run had ultimately been removed as a result. It should be noted that the moderators had conducted the statistical analysis, and had compared Dream’s live stream of the speedrun and 6 other consecutive runs by him, as well as 6 other live streams by other well known speedrunners to produce the odds mentioned above.

Dream had retaliated by accusing them of misusing the data to produce odds that would look bad to him, claiming the analysis was constructing from false odds just to ruin his reputation. He attempted to disorove the odds given by the moderators by hiring a astrophysicist from Harvard to conduct an analysis with the data (as well as critique the attempt by the moderators) and find that the odds were within 1 in millions. In fact, he even further analyzed that the Binomial Distribution could not be applied to find this chance, as a third value from true/false can happen.

**My Hypothesis and Analysis Plan**: I will first be conducting an analysis of both of these two, and decide which one had used more applicable and more authentic results, while figuring out any flaws the moderators and the astrophysicist had conducted in their analysis. Furthermore, I will try to use standard analysis with the data provided by the moderators.

Finally, I will explain and conduct rebuttals for any statements and contradictions each side may have made in their Report. Please note this is a different form of hypothesis test many statisticians conduct themselves with, as evidence is more widely used in this report and the reported incident in general.

H0: Dream did not cheat at all, we just observed very random events happening.

H1: Dream had used an illegal client to increase his odds, thereby creating his own luck and cheating.

**SIDE A: THE MODERATORS**

# Here is the dataset of Dream, the main data

#Note: p represents Ender Pearl Data, b represents Blaze data Column A represents the amount of gold traded in that run, Column B represents the pearls recieved. Note that in this version, Pearls have a 5% chance of dropping for any unmodified version of Minecraft.  
library(readxl)  
dream\_p <- read\_excel("Bartering stats.xlsx",   
 sheet = "Dream", range = "A2:B24")  
dream\_b <- read\_excel("Bartering stats.xlsx",   
 sheet = "Dream", range = "A2:B35")

#Here we will also get the data of the runs of the four other members that the moderators compared this data against

runner1\_p <- read\_excel("Bartering stats.xlsx",   
 sheet = "Benex", range = "A2:B22")  
runner1\_b <- read\_excel("Bartering stats.xlsx",   
 sheet = "Benex", range = "A2:B13")  
runner2\_p <- read\_excel("Bartering stats.xlsx",   
 sheet = "Illumina", range = "A2:B23")  
runner2\_b <- read\_excel("Blaze stats.xlsx",   
 sheet = "Illumina", range = "A2:B17")  
runner3\_p <- read\_excel("Bartering stats.xlsx",   
 sheet = "Sizzler", range = "A2:B25")  
runner3\_b <- read\_excel("Blaze stats.xlsx",   
 sheet = "Sizzler", range = "A2:B18")  
runner4\_p <- read\_excel("Bartering stats.xlsx",   
 sheet = "Vadikus007", range = "A2:B26")  
runner4\_b <- Blaze\_stats <- read\_excel("Blaze stats.xlsx",   
 sheet = "Vadikus007", range = "A2:B12")

Here is a plot of the Cumulative distributions between Dream’s trades and the other Runner’s Trades:

In the Cumilitative Graph Shown, One thing looked out of place, and brought the most attention to many users: Dream’s curve. While the graph is not enough to make a conclusion, suspicion comes when you see that while 3 other speedrunners maintained a low variance in terms of their data, Dream has a high variance to all 3 curves.

**Overall, Three Values were presented by the moderators as to the odds found in the 2 datasets that Dream had “Been lucky”**. Those values are:

or 1 in 8 Billion, for having such luck in terms of the 6 consecutive Streams for collecting Ender Pearls in his 262 trades where he got 42 Ender Pearl Trades.

or 1 in 113 Billion, for having such luck in terms of his 6 streams that he collected 211 Blaze Rods from 305 Blazes.

or 1 in 7 Trillion, for having the luck of both events occuring.

How we will get to these values will be explained below. We will be incorporating the **Binomial** Distribution to explain these values, but an explanation on probabilities must be conducted first.

**Explanation**: into the Mechanics

A more in depth into the probability on **Pearl Trading** is this. Similar to a prize, or a sort of lootbox, trading a gold item to a NPC known as a Piglin yields different items. All of the items are irrelevant to what the speedrunner mainly wants, except for 1 item, Ender Pearls. Each item has a weight/percentage associated to it. Ender pearls trades have a weight of or 5(4.7)%, and each trade can yield a random number of enderpearls. Overall, at least 10 enderpearls are needed to finish and stop trading.

This means alot of time is needed to be dedicated during a speedrun on trading alone, which is why a binomial distribution can be done on this aspect along with blaze rod drops.

Blaze rod drops are very simple, a hostile mob called a Blaze spawns, and players must kill them for a drop that they have a chance of releasing called **Blaze Rods**. Blaze rods are the only item that can drop from a Blaze and has a 50% chance of dropping upon death. players need at least 5-8 rods, then can stop collecting.

Overall. a Binomial distribution is perfect with terms such as these. You can either acquire Ender pearls via the trading system vs failing to acquire them (such as receiving a random item from the trade that is not Enderpearls) or get a Blaze Rod vs not one dropping from a Blaze. Either **Success or Failure**.

A reminder on the Binomial Distribution formula is below:

where

represents the number of total trials. In our case, it is the number of times in the dataset a Blaze had been killed or a Trade was conducted

represents the success numbers. In our case, it is the number of times a Blaze had dropped a Blaze Rod or a trade yielded Ender Pearls

represents the probability of success from 0 to 1. In our case, either 0.047 for the probability of a Pearl Trade or 0.5 for a Blaze Rod dropping.

Applied Below is the computation

In mind of the Binomial Distribution being very unfair and looking like a possible attempt by the moderators to used biased data, the moderators attempted to correct for the probabilities mainly in two ways.

One of them was using the Cumulative Distribution formula as shown below:

$$

$$ As this can account for the probability someone can get anywhere from to , rather than the exact number of 42 in 262. This gives a bigger probability.

The other way they accounted for bias was via accounting for a stopping rule the data posed. Say Dream was never cheating, and the data collected all this time was just from luck and at the point that the data stopped was the point where Dream’s luck ran out. This would make the data seem very malicious and unfair to the true probabilities.

As a result, a stopping rule algorithm was conducted. Rather than account for the probability that someone can collect ender pearl trades, the stopping rule algorithm tests the chance that someone can get the probability from the trades at any time during several more streams that were conducted. Initially, 6 streams were tested to this, which was done by the moderators to provide a wider curve which can strengthen the probabilities. It would be similar to looking for an event where you get a 4% or more chance event to happen across 50 different occurrences.

**Translation**: The analysis turned from them trying to find the probability of the 42 Ender Pearl Trades to finding out what is the chance Dream would get a 1 in 82 Billion chance across all of his streams.

I will not be able to provide an exact calculation with this algorithim. This is mainly because it was used through Javascript, and while the code is listed on their report, it requires a large amount of interpretation of each line that I cannot provide. Therefore, I will be going over their result from using that algorithm.

Conducting the Cumulative Distribution Formula for Ender Pearls Yielded the following: For Pearls

For Rods:

and combining the chance of both occurring in the **EXACT** set time frame is simply a matter of multiplying the two values.

(5.65\*10^(-12))\*(8.79\*10^-12)

## [1] 4.96635e-23

You might notice these values are different from the three values previously mentioned. This is because the values here were felt to be biased to the moderators. As a result, the moderators used the stopping algorithm and setting this probability across 11 livestreams which Dream did during his speedrunning sessions, compared to the 6 examined in the dataset. Using their algorithm for stopping, they found a value of . the next step came in bounding the probability in terms of it coming across 11 livestreams, AKA checking if that probability mentioned would happen at all within 11 livestreams where he did the same exact thing.

This led to the following odds.

This probability meant to represent the probability what happened in the dataset could ever occur again. It was even made more generous by powering the following possibility by 1000, so as to include the fact more than a 1000 speedrunners do speedruns on minecraft.

This led to a final probability value for pearls as or a 1 in 8 billion chance.

Blaze Rods, with the same examination, led to a value of AKA the 1 in 113 billion

Finally, combining these two led to utilizing Fisher’s Method, AKA taking -2 times the sum of logs of the p values. and determining the CDF. Leading to a value of

this, estimating along with accounting for investigation bias by accounting for 90 possible options rng may come to play throughout speedrunning come in, multiply the probability by 90 and come to a value of

or the 1 in 7.5 trillion chance.

**SIDE B: ASTROPHYSISIST AND DREAM**

After several days, Dream hired his astrophysicist to make a fair, neutral look at the datasets and probabilities and construct his own perspective. Regardless of the outcome, it would be presented whether or not the results were favorable to him.

To start, the Astrophysicist made a 19 page rebuttal on the data set. The core elements described were this. He and Dream felt that there were too many unknowns that were not factored. For one, he argued that each Minecraft speedrun is set in a different world seed, which means different events can happen. He had tied this to the weight of Enderpearls and Blaze Rod chances, explaining that with each new seed, there may be a chance of both events occurring, but didn’t provide proof.

Additionally, he chastised the moderators for using the Binomial Distribution, ultimately deciding the Bayesian form is more suitable due to the third event “Unknown” happening within the data collected. This would mostly state to be times where the camera did not record each trade. In addition. He noted that instead of there being 90 random events that could account for the probability, he stated over 1000 probabilities could occur. That is a massive exaggeration, as most times last less than 10 minutes, meaning that is a probability happening almost every 10 seconds.

He had used the likelihood estimates in order to find the probability for Ender Pearls to be a whole ten times smaller than the previous! He had admitted though that with his setup, it was not the best fit when including a stopping criterion of his own.

He found for the Blaze Rods, with the Bayesian Probability to be

Ultimately, he found the chance to be 1 in 10 million for both events to occur., citing the 1 in 7.5 trillion chance is too extreme.

**My Analysis**

Either way, we can reject . Both probabilities presented by the moderators show that the something more than randomness may possibly be at play. Both are smaller than 0.05, which is considered the significance level. Therefore, with a probability that Dream had odds smaller than 0.05 that he randomly encountered these odds, without any altercation to his system, we cannot say that this probability is likely, and we reject this chance in our analysis. Instead, we will consider the alternate Hypothesis.

For Dream’s side, several inconsistencies are noticeable. Let me just conclude though with one confirmation. Dream cheated, he cannot naturally get odds as high as 1 in 7.5 trillion. those odds are realistically hard to reach in just any century for that matter.

For the moderator’s analysis, I do find their work very authentic. For the Astrophysicist that was hired by Dream, he had examined the data, and instead of copying or making similar works of the moderating team, he had decided on his own to make a rebuttal on the data set and describe the flaws he claimed to find. While I don’t agree with many of his sentiments, several things are true.

Firstly, the moderators made very sound arguments, one of their finest arguments being the utilization of a stopping rule rather than simply taking one occurrence and testing that probability by itself. However, by conducting a specialized algorithm, and by going very far and sophisticated into their data collection and generosity to avoid bias, they failed to fully explain it to those wanting to understand it. While things like statistics is suitable for college students and high school students, the explanations in their pdf become unclear to everyone else. The work done by the moderators was presented by the public, most of the core audience being video game players who were not well established in basic statistics.

Second, the moderators had found the correct probability for the Enderpearl trades and Blaze rod drop chances, and found across the several live streams tested for Dream. All of the math associated in their document is correct. I agree with this sentiment. Alongside my agreement with the data came the astrophysicist’s which is why I believe he decided to argue against the data itself and the compensation for the many unknowns that the moderators took to limit the probability to be more generous for Dream.

One of Dream’s rebuttal that I believe can be considered deeply flawed is their claim that the binomial distribution cannot work because there were around 40 unknown trades not listed due to the livestream not looking at every trade, so therefore some trades not listed must be categorized as “unknown”, leaving true, false, and unknown as events. While I do see n+40 being very important to include, **That is more than 10% of the data!** This of course would prevent the BDF from being used if true. However, unknowns are not what the data is being examined for. It is considered integral to know **whether or not the trade was a pearl or not**. Placing unknown outcomes into a Binomial Distribution, or any distribution whatsoever is a flawed practice which serves to muddy the data. I believe the moderators took the best decision and simply removed the unknown outcomes for consideration rather than keep them.

The astrophysicist, while making a valient attempt, does not recognize Minecraft mechanics enough compared to the Moderators, who spent thousands of hours on the game to know its in depth mechanics. Additionally, there exists 2 noticible errors within the document in regards to calculations. they experimented with an example done on coin flipping within the Moderator’s document, getting 20 heads in 100 coin tosses. Apparently, there was no calculation for this present, and the value is noticeably wrong.

Conducting my own analysis and comparing them to others, they and I concluded that the astrophysicist simulated the odds for heads and tails incorrectly. I predict that when factoring for heads they had divided the denominator by 4 instead of multiplying. Which is why it is 4 times off of the true value,

In the second error, from their use of the bayesian model, while possible for this dataset, may have muddied the probabilities too. there is no need to explain that there is only two possibilities that should be occuring: getting the blaze/pearl vs not getting. In fact, utilizing likelihood as well for the Pearl’s side may have contributed to him getting very very different values. It should be noted that as to this point, I do think the Moderators were also very lenient in making probabilities very generous in their attempted correction of Bias.

When finding a probability of 1 in 7.5 trillion, that is odds people cannot come to even if they have 1,000,000 people doing the attempted action for those odds several hundred times a day.

In general, What I learned from this is that with statistics, you can generally determine if an outcome has happened or not, even despite only seeing a property. While the moderators could not go back to Dream’s computer and detect any false modifications, they looked at the dataset, provided a probability with less bias, found values that are lower than hitting the lottery 7 times, and made a inference using it. You do not need hard evidence, you can use statistics to draw conclusions. Although while it can’t be perfect for everything, statistics can help guide you where to look in terms of data searching.

**--------------------**Lastly, a note of events that happened after.**------------------**

**Clay had found out a mod of his had been tampered with by a friendly plugin designer of his**.

He, unknowingly, brought a mod which was generally allowed in Minecraft speedrunning. However, the mod had increased drop chances for both Pearls, and Blaze rods.

He confessed to this act later down the line after finding out from the designer many months later. Although his speedruns are taken down, the other ones which were found not to be cheating remained and he continues speedrunning to this day.

Do I believe that Clay was lying? No. For one, the fact that Clay had brought this up rather than not reveal it brings a tiny amount of credibility. Secondly, in his reasoning for why the mod had these, he had claimed that it was for a minigame that he does with his friends in Minecraft, where he boosts the drop chances for both in an effort to speed up the game. Coincidentally, this minigame involved speedrunning. So it was possible the mod creator simply made things easier for Dream not realizing he would add the mod in while doing a speedrun, or he could have forgotten about the drop chances being added in the mod.

Knowing what I know now, would I change my research? Yes. I would try to look at the data to see where the signs showed, and see now how I myself can see if there is any signs of cheating from looking at the datasets alone.

**Data Sources and Reports**:

<https://mcspeedrun.com/dream.pdf>- Moderators Report

An excel spreadsheet of Dream (as well as other speed runners named Benex, Illumina, Sizzler, and Vadikus007)’s Piglin Bartering/Blaze rod chances. Provided by moderators from the speed runners videos. The observation numbers fluctuated, but they were taken from around 6 livestreams. <https://docs.google.com/spreadsheets/d/1NJTdZnkF10nw2tDIS5hZZx8KmC2PC6I71XGtzc5iXLE/edit#gid=0>

<https://docs.google.com/spreadsheets/d/1P58S94yKB3Bm4A4_VotWyeelk_PvaTE1nDZx9DalEyk/edit#gid=1130044964>

• Dream’s astrophysicist’s statistical findings/counterargument. PDF form, contains several calculations done with the data and finding above: <https://drive.google.com/file/d/1yfLURFdDhMfrvI2cFMdYM8f_M_IRoAlM/view>