Minecraft Xbox One Statistics

Applying statistics on a videogame to find interesting things

Evan Tyler

School of Business, Stockton University

Probability and Applied Statistics

May 2, 2022

**Minecraft Video Game Statistics Xbox One**

Evan Tyler

Minecraft is a massive game with almost endless possibilities. I have played it since around the 2012-13 era. So, it is safe to say that I know and have achieved a lot in this game as time has progressed onward. Now, there really isn't a clear beginning and end to Minecraft as it is a survival sandbox game. Since a sandbox game gives you free reign on what you want to do, we will define originally what was the beginning and what was the end. The beginning of the game has you spawn and explore and one of the first things a lot of players do is open their inventories since an inventory is needed to see the player themselves and to store items. The end of the base game is when you defeat the final boss, the ender dragon in the end. You get an achievement for defeating the final boss which isn't the final achievement but for this paper we will define it as the end and the last achievement. Now there is debate on how long it takes to beat the base game content. But from the website I used it determined that the total time to log in to beating the ender dragon takes around 35hrs. I will use the information gathered about achievements and in-game events from sources listed below to apply techniques learned in class to find interesting things out about the game and its achievements. If you don’t know much about Minecraft and are confused by some of the things, I talked about I link the Minecraft Wiki as a source down below. I use the Minecraft wiki throughout this paper to gather data to use to make certain problems. However, it can be used a reference for things not understood mostly Minecraft lingo.

Chart, bar chart

Description automatically generated

The first interesting thing we can do with this player achievement data is see how many achievements have been unlocked by the total number of players. I did this by grabbing data from the pages the site had of individual players who had played the game. We can do this by creating a frequency distribution which will tell us how many achievements in a range of 11 for each interval number of players unlocked. In the beginning of the frequency distribution, we see a massive spike which makes sense as a lot of players get easy achievements and these stack up as the player is fully interested in the game. A trend forms which is a downtrend from 0-11 to around 84-95. This can be explained since over time most casual players lose interest in the game and may logout before gaining more achievements. However, that could also be the opposite case as players could play the game for a long time and be focused on other things then gaining achievements. Since like I explained before earlier on that Minecraft isn’t a linear game and is a sandbox. Which means players can do whatever they want and not have to progress or earn achievements. Which combined with other players losing interest adds to the downtrend witnessed in the frequency distribution. Now the spike in the end of the frequency distribution from 6-106 to 107 - 118 also makes sense. Since the goal of the website is to track your statistics in games and compare it globally with others it would make sense many of the players who are registered to the site are invested in achievement gathering. Which explains our spike in the last interval in the distribution meaning a lot of the players on the site have gathered most if not all of the achievements in the game.

Now Minecraft is infamous for its large content updates which bring large amounts of content to the game. For example, just recently the game got the caves and cliffs update which added new mobs and content to the base game. Now we can use these updates to the base game as an example for sets. Sets are a collection of objects that form a group. In this case we can call a set A which includes all achievements in the game in which there are currently 118 current achievements. This would be the universal set since this set would include all achievements in the game. This would be listed as set A = {1,2,3,4…..118}. Now, there are updates that come out almost every year for Minecraft adding a large new content update and with this update comes more achievements. So, we can have a subset B which includes achievements from the Update 1.17.0: Caves and Cliffs part 1 which includes B = {111, 113, 114}. Since the update only came with 3 new achievements. Now for this to be a subset the superset needs to have subset B included plus more elements in which this case set A has the elements from 1-118 which is more than just subset B.

Permutation and combinations are another important topic covered in this class. Combinations can tell us the total number of things arranged when the order doesn’t matter. For example, say you’re designing a house in Minecraft, and you want to build a wood pattern out of planks. Making this pattern would be placing two plank blocks next to each other. This pattern can have 2 planks out of the 8 total planks available in the game currently. So, we can use the combination formula to find out how many ways the 2 pattern planks can be organized.   = = 28 which this tells us that there are 28 different ways in which the two-plank block pattern can be laid out. Since the order doesn’t matter in that case since, we just wanted to know how many different combinations there were of the two-block pattern. Now there is also another important topic covered which is permutations which tell you the total ways that something can be arranged however order now matters. Say, there was a speed running competition to beat Minecraft which this would be signified by getting the “The End” achievement. Now say in our theoretical contest there is a first, second, and third prize which is going to be awarded to 30 speed runners who have entered our competition. We want to find out the total ways in which the first, second, and third prizes can be awarded to the 30 speed runners. We can use permutation for this since order matters. = = 24360 different ways the first, second, and third prize can be awarded to our 30 speed runners in our competition to earn the last achievement and beat the game.

Now there’s some other interesting topics that we did cover that can show us some more interesting data about our own data. Such as the Multiplicative Law of Probability which can tell us the probability of the intersection of two events or getting a and b. This formula can be used here if the events are not independent 𝑃(𝐴∩𝐵) =𝑃(𝐴)𝑃(𝐵|𝐴) now if the events are independent, you can use this formula 𝑃(𝐴∩𝐵) =𝑃(𝐴)𝑃(𝐵). Now say we want to calculate the probability that a player has both a new and old achievement. We classify old as being an achievement from the base game and a new achievement being from the newest update. Our new achievement will be “Sound of Music” and our old achievement will be “MOAR Tools”. Now these events are independent since getting the “MOAR Tools” achievement wont effect getting the “Sound of Music” achievement or any achievement in the game. So, we will use this formula 𝑃(𝐴∩𝐵) =𝑃(𝐴)𝑃(𝐵) Now the prob of the old achievement “MOAR Tools” is A = 64% and the new achievement “Sound of Music” is B = 3%. Now we plug in 𝑃(Player getting new achievement∩Player getting old achievement) =𝑃(.03)𝑃(.64) = 0.0192 which can be rounded to ~ 2% of players have one of the old and new achievements. Which makes sense since this update came out recently so the probability of having both would be extremely low. Now another law we learned in the class was The Additive law of probability which tells you the probability of the union of two events or getting A or B. Now we can apply this to our data on our achievement tracking website. This can be done by using this formula here 𝑃(𝐴∪𝐵) =𝑃(𝐴) + 𝑃(𝐵) − 𝑃(𝐴∩𝐵) now let’s say from the multiplicative problem we just did we wanted to find the probability on the site that the player had either the old achievement or the new achievement. Again, these achievements are independent events since unlocking an achievement doesn’t affect the outcome of unlocking the next achievement. So, we can use the formula provided above. 𝑃 (Old Achievement ∪ New Achievement) =𝑃(.64) + 𝑃(.03) − 𝑃(.64∩.03) = .67 - 𝑃(0.000192) = .669808 = ~ 67% that the players have one or the other achievement.

Another important topic that was discussed in this course was conditional probability which in a sense means what is the probability of an event occurring given another event has already occurred. In this case the probability of A given that the B has already occurred. We can solve something interesting here with conditional probability. Now, an interesting thing to figure out would be the likelihood that a player has “The End” achievement from beating the final boss while also having the “passing the time” achievement? This is interesting since the website true achievements tracks the average completion of the game in 30–35-hour interval. Now this is interesting because we can calculate the probability of a player beating the dragon if they have 100 in-game days which in Minecraft one day = 20 mins so 100 days would equal around 33.333 hours in real life. To do this we need the percentage of players getting both achievements “Passing the Time” is owned by 41% of players while “The End” is owned by 21% of players. Now we plug this into our formula A would be 21% and B would be 41% = 21% now this means that since it is the same as the probability of defeating the dragon initially that the events are independent since the probability of A given B didn’t change the probability of A. So interestingly enough having the “Passing the Time” achievement doesn’t affect the player from having the “The End” achievement.

Baye’s Theorem is also another important theorem since it might seem similar once the formula is given since it solves the probability of a given b or b given a. However, there is a clear difference between just plain old conditional and Baye’s theorem which makes it a powerful theorem. This difference being the lets say your out in the field and gather a data set and you can’t figure out what a given b but b given a is much easier to calculate well you can use Baye’s theorem to calculate it using b given a. Again, sometimes calculating both a given b and b given a is indeed easy but in case it’s not Baye’s theorem can take the probability that is easier to get and find the harder conditional probability from the easier one. Both conditional probabilities are related by Baye’s theorem. Again, I will demonstrate that now with my dataset. Say that we want to find out the probability of getting the achievement for beating the game which we have used before called “The End” but we want to figure out how many users have defeated the wither (which is another boss in the game) and have attained the achievement for beating the wither which is called “The Beginning” before they have attained “The End” achievement. In this case “The End” achievement would be A and is owned by 21% of gamers and “The Beginning” Achievement would be B which is owned by 16% of gamers. Say in our case getting that conditional probability of A given B is difficult so, we can use Baye’s theorem instead to use the other conditional probability B given A to find A given B. Let’s look at an example using the formula first we have to calculate = = = 16% which makes sense since getting the wither and ender dragon achievements are independent getting the ender dragon achievement before the wither has no effect on the probability of B since it is independent. Now we can plug it back into the main formula = = = = 21% which again makes sense since the achievements are independent we would expect getting the wither achievement “The Beginning” to have no effect on the probability of getting the ender dragon achievement “The End”. Now this shows that Baye’s theorem can be used to get one conditional probability if the other is known. Which again is useful because one might be more difficult to calculate then the other.

Now the next section or sections will get into the discrete probabilities that we learned over the course of the semester. The first distribution I want to introduce to the dataset would be the Geometric distribution. Which can help if per say we wanted to know on a specific trial say randomly selecting gamers account on the website would be the probability the 7th gamer was the firt one to have a certain achievement? Well lets figure it out the achievement “Return to Sender” has a 27% chance of being in a gamers account. We want to find out out of 10 randomly selected gamers what is the probability the 7th gamer have the “Return to Sender” achievement? The Geometric distribution formula is as follows = p(7) this number equates to 4.0% probability the 7th gamer has the “Return to Sender” achievement. Now we can also calculate the average or the expected value which is = . = 3.7% So that means the average value to get a first success would be around 3.7 which we could round up to the 4th trial. Now for the variance now if we plug in which this measures the spread between values in our data set. Now for the standard deviation 3.16.

The binomial distribution can calculate x successes on n trials being that the trials are independent of one another which also have two outcomes usually success and failure. Let’s use this in practice in our dataset now the wither boss achievement “The Beginning” the probability a randomly selected gamer has the ‘The Beginning” achievement is 0.0085%. If we pick 20 randomly selected gamers registered to the site that what is the probability 5 of them have the wither achievement? So, to use the binomial distribution, there must be two outcomes in this case the success being the gamer has the achievement and the failure being the gamer doesn’t have the achievement. So, in this case = 6.05E-7 which is an extremely small percentage. Which makes sense as it is a very small percentage a randomly selected player has the achievement in the first place and almost a 99% failure rate. Now, that is not all we can calculate the mean which is called the expected, the variance, and standard deviation. The Binomial distributions expected formula is: again this refers to the expected from our problem that we just did above with the wither achievement. = .17 which means there’s an average of .17 that a player has the wither achievement which makes sense as it is an extremely low percentage in the first place. Now for the variance = 20 \* .0085 \* .9915 = .168555 this measures the spread between values in our data set. Now for the Standard deviation in which the value itself is the average of how far values are away from the expected value. The value of standard deviation is derived from the variance and that is shown here since it is only the square root of the variance ~ .4106.

Another distribution that we learned was the Poisson distribution this discrete distribution can tell us how likely the amount an event is likely to occur over a period of time. Now for an example say we wanted to figure out wheat in Minecraft update once (grows) at an average of around 1 tick per 68 seconds the number of ticks will follow a Poisson distribution. Now what if in a randomly selected 140 second period there are exactly 2 ticks which equates to 2 updates? We can use the formula to find out = = % means that the chance of having the wheat update twice in a period of 140 seconds is extremely low. Meaning the wheat would most likely take longer than 140 seconds to update. Now we can calculate the mean and the variance which the formulas for both are: 𝜇 =𝐸(𝑌) =𝜆 and 𝜎2 =𝑉(𝑌) =𝜆 which both end up equaling 𝜆 so the mean and the variance end up both = 140. Now the standard deviation is equal to the ~11.8.

Now another interesting theorem Tchebysheffs theorem helps us in estimating least proportion of operations that are within a specific number of standard deviations away from the mean. Now lets solve a problem say you had and average completion rate of the game being 35hrs with a standard deviation of 10 hours what would be the minimum percentage of hrs it took to complete the game between 20 – 50hrs? We know this is a Tchebysheffs problem since we can check that 35hrs is between our interval of 20 and 50. In this case it is since it is 15 away from the low and top interval. It also doesn’t tell us the shape of the distribution which is fie since the Tchebysheffs theorem can be used on any distribution. To solve this problem we can do it like this = = 1.5 to check that this was the right approach we can plug in the lower limit and solve also if we were correct the answer should be the same but with an opposite sign = = -1.5 this again tells us since they qual eachother besides the sign that it is a good interval to use Tchebysheffs theorem with. Now to solve the rest of the problem the theorem states that at least (\*100% of the data will lie inside the given interval = at least = ~ 55.5% which means at least 55.5% of players who finish the game will be inside the interval of 20 – 50 hrs.

The uniform distribution is another distribution we learned the formula is as shown here

Now, in class we learned a simpler way to solve the uniform distribution in which I will be showing off here. A uniform distribution problem will have a range from a to be in this case then we will have to solve to see if y is above or below the uniform distribution in between the two values. We can do this by making a number line which will be shown as the problem is done. Now for the problem, say from looking at our chart on the first page of the report we see that in 3.5 hours which is the average amount of time new players will log on and quit the game the uniform distribution of achievements earned in our first bar is 0-11 say we want to find out the probability the next player that logs on and quits will have > 8 achievements unlocked. Well lets find out now are interval is (0 to 11) so we can create a number line 0🡨-|(1)-|(2)-|(3)-|(4)-|(5)-(6)-|(7)-|(8)-|(9)-|(10)-🡪. Now in order to find the probability that the next gamer unlocks 8 achievements you count the spaces in between the values >8 so there is a space between 8-9, 9-10,10-11 therefore there is a 3/11 = 27% chance that the next gamer who logs out in the 3.5 hour mark will have > 8 achievements.

I also wanted to add an extra paragraph to vouch for extra credit. I wrote 6 pages minus the chart which exceeded the minimum of 2 pages. These are also single spaced which equate to 12 pages double spaced. I included a problem from every single chapter in the textbook which we went over in class. Including multiple topics from each chapter in most cases. I spent days on days looking up the statistics in Minecraft and coming up with different interesting problems to get interesting information. In my formula sheet I added every formula from every chapter covered. Plus, bonus formulas that aren’t even formulas that we covered in class from chapter 5. Which caused me to write and entire extra page for the formula sheet which took a bunch of extra time. Plus, on my Jfree and Commons program I did the plot, salt, and smoother with my old code just using JFree chart with incrementing x values. Plus, I did the project with the combo of JFree Chart and commons with random x vals to cause more extreme salting and smoothing. I did all this extra work while being out with the Flu for an entire week during the end of the semester which made this project much more of a crunch. However, I still persevered, and I believe still put in a great amount of my time and effort into this report and project.

References

“Minecraft Achievements.” *TrueAchievements*, https://www.trueachievements.com/game/Minecraft/achievements.

“Minecraft Wiki.” *Minecraft Wiki – The Ultimate Resource for Minecraft*, https://minecraft.fandom.com/wiki/Minecraft\_Wiki.