Geometric Distributions & The Birthday Paradox: Crash Course Statistics #17

Crash Course: Statistics
https://youtube.com/watch?v=5VMTeBoEcQg
https://nerdfighteria.info/v/5VMTeBoEcQg

Hi, I'm Adriene Hill, and Welcome back to Crash Course, Statistics.

We have to wait for a lot of things in life. We wait until we're old enough to live on our own, or go to college, or drive a car.

Waiting can suck, and it's even worse when you don't know how long you'll have to wait. Luckily, in certain situations, probabilities can help you guess how long it might take for something to happen, like getting a full house in poker, having your first daughter, or winning the lottery. INTRO For example, you're eating from a box of Bertie Bott's Every Flavour Beans during a fun, if risky, hang out with your friends.

These Jelly Beans have some flavors that are awesome like Cherry and Peppermint, and some not so awesome like Grass, or Boogers, or Vomit. Ewww. The problem: you don't know if you're gonna get a good or gross flavor until you eat it.

And you know that your affinity for this game will go away if you're ever unlucky enough to come across one of those Vomit flavored beans. Cinnamon! It's delicious.

But how likely is it that you'll be able to eat 4 of these odd Jelly Beans before you get that dreaded Vomit flavor and decide to get new friends? Turns out, there's a formula to do this. The Geometric Probability Formula.

This formula comes from The Geometric Probability Distribution, which looks similar to the Binomial Probability Distribution that we talked about in the last episode. But they do something a little bit different. Geometric probabilities tell you the probability that your first success will be on your nth try.

And success here just means the event we're interested in happens, it's not always a good thing. For example, the probability that your first Vomit flavored jelly bean will be your 5th Jelly Bean this means that the first 4 beans were all not vomit flavored. Let's pretend the probability of getting a Vomit flavored bean is 5%, then there's a 95% chance that you'll get any other flavor.

As you remember from the multiplication rule. Our chance of getting four non-vomit beans in a row is 0.95 x .95 x .95 x .95 or .95^4. The probability of then getting the Vomit bean is 0.05, so all together, the probability of getting 4 non-vomit beans and then 1 Vomit flavored bean is this or about 4.07% In a more general form, The Geometric Probability Formula says that the probability of the kth try being your first success, is the probability of failure to the k-1 power, times the probability of success to the first power.

Which is what we just did in the vomit jellybean example.. The probability of the 5th try being your first success was the probability of 5-1 or 4 failures and then one success. If we graph the geometric probability for all possible values of k, we get a Geometric distribution that shows us the probability of each trial being the first time we get a success.

For example let's look at a section of the geometric distribution for finding the probability that each trial is the first time we'd eat a vomit flavored jelly bean. You'll notice that I said "a section of" because the geometric distribution could go on forever technically we could eat thousands of Bertie's Beans and still not find any Vomit Flavored ones...in fact that's pretty much what I hope when I eat them. It's just very unlikely, which we can see on our graph.

As k increases, the probability of that being your first success gets incredibly low, mostly because you'll probably have found the flavor that shall not be named already. Let's look at another example: You're finishing up basketball practice and your coach announces that you'll have to make one free throw before you can leave. You're really bad at free throws, and you have a 20% probability of making any given shot.

So we can calculate the probability that the first shot you make is your 10th shot. Plugging everything into our formula, we discover that the probability of this particular scenario is this: or about 2.7%. The Geometric distribution for this example looks like this which gives us all the probabilities that a certain trial is the first time you make a shot: But you'd also get to leave if you made it before your 10th shot.

To calculate the probability that you'll have to shoot 10 or fewer free throws before you make one, we can add up all the geometric probabilities from 1 up to 10. This essentially tells us the probability that it takes k or fewer tries before your first success which you might remember from when we talked about cumulative distributions. A cumulative geometric distribution is incredibly useful because often when we ask questions about how long we have to wait for a certain outcome, we want to consider whether it happens before the nth trial.

Like we said, whether it's the first or 10th shot, you still get to leave practice, and it turns out that there's a 89.3% chance that before your 10th shot, You will have made a basket. That make for a total of about 92% chance that you'd make it on or before your 10th shot. If you want to find the average number of shots you need to take before you score, you can use the geometric distribution to calculate its mean, which is 1 divided by the probability of success or this.

The mean number of shots we'd need to take before we made our first would be 1/0.2, which is 1/(1/5) or 5. And this makes some sense. The smaller the probability of success is, the more tries we'll need--on average--before we get a success.

For example, if there's only a 10% chance that we'd make the shot, we're really bad basketball players we'd now need to try an average of 1/(1/10) or 10 throws before we sink that perfect shot. The mean and cumulative frequency of the geometric distribution can help you weigh your options. Say you're at Target, and you see the display of Pokemon cards.

You really want that Pikachu card, but let's say that the probability of getting one is 1/200. You have enough money to either buy 4 packs of cards, or Star Wars Episode VI on Blu-Ray. Star Wars guarantees ewoks.

Pokemon means maybe Pikachu. You're really into small cute animals...small cute animals with Static ability...even better. You want that Pikachu.

Is is worth it? If each pack of Pokemon cards has 10 cards, you'll have 40 cards, each with a 1/200 chance of being a coveted Pikachu. The probability that you wouldn't get a Pikachu until your 40th card is only about .04 %, but the cumulative probability that you'd get a Pikachu on your 40th card or earlier is about 18%, which isn't bad.

So now you're left to decide whether it's worth it to forgo an evening with ewoks for an 18% shot at getting your beloved Pikachu. That part is something statistics can't decide for you. Before we wrap up probability...let's look a really fun statistical paradox...the birthday paradox.

There are 365 birthdays that anyone can have, we're not going to count leap days-- to keep things simple. So let's say there are 20 people in your classroom. What's the likelihood that there would be any shared birthdays?

To me...it seemed like it'd be pretty low. Let's assume that there's an equal probability of having a birthday on each day of the year. That's not quite true, but it's close enough.

The first person we look at has a 100% chance of having a unique birthday, simply because we haven't looked at anyone else yet. The second person also has a pretty high chance of having a unique birthday, they could be born on any of the 364 days on which person #1 wasn't born, so there's a 364/365 chance of having a unique birthday. Similarly person #3 can be born on any of the 363 days on which neither person #1 or person #2 are born, so there's a 363/365 chance that they'll have a unique birthday as well.

And this pattern continues for all the people in the room. By the 20th person, there is a 346/365 chance that they have a unique birthday. Using the multiplication rule, we know that the probability of ALL of these events happening--in other words each person having a unique birthday-- is the product of all these probabilities.

So the probability that everyone has a unique birthday is about 59%. That means that the other 41% of the time at least two people will share a birthday. Once you get up to a group of 70 there's a 99.9% chance of someone will be sharing a birthday with someone else.

This is one of those instances where statistics trumps my own intuition. And a good reason to be prepared for the possibility of double cupcakes if you have a big class. Probabilities are important.

They're things we use all the time. When you're at your local diner and see those super cool Crane machines, you might estimate the probability that you can snag that sweet stuffed otter by the time your food arrives and whether it's worth your time and money to even try. Or you can use them to figure out that on average, you'll probably have to wait a looooong time before you win the California State Lottery , like a really long time, on average it should take you about 15 million tries.

That's a lot o' lotto tickets. Brandon made me say that. Geometric probabilities, and probabilities in general, allow you to guess how long you'll have to wait for something, so you can decide whether it's worth it.

As the famous mathematician Pierre-Simon Laplace once said, probability "is basically just common sense reduced to calculus; it makes one appreciate with exactness that which accurate minds feel with a sort of instinct, often without being able to account for it." In other words, it allows us to quantify things we already feel. Probability assigns numbers to common sense. And we all need a little more of that.

Thanks for watching, I'll see you next time.