The Binomial Distribution: Crash Course Statistics #15

Crash Course: Statistics
https://youtube.com/watch?v=WR0nMTr6uOo
https://nerdfighteria.info/v/WR0nMTr6uOo

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

When you're using the patterns of probability to predict events in your life, it can be nice to have some mathematical shortcuts. That way, when you're reading People Magazine, you won't have to spend so much time sidetracked, calculating how likely it is that Harry and Meghan will have 3 boys and 2 girls if they have 5 kids.

And you won't have to pause Willy Wonka and the Chocolate Factory, right as Grandpa Joe starts singing, to figure out how likely it is that you could have found one Golden Ticket by opening 4 chocolate bars. Today, we'll introduce you to some of these shortcuts so you can get back to the movie. Before we start today, heads up -- these shortcuts are equations.

And we try in this series not to spend too much time plugging numbers into equations, but today we felt like we kinda needed to. Because understanding these things is important. So to sweeten the deal, we added some zombies.

INTRO You're in your kitchen, making a piece of toast with your old toaster--which has seen better days. You've figured out that each time you make a piece of toast you have a 20% chance of being shocked--not lethally, but painfully. Alright so maybe it's time to get another toaster.

But you haven't had a chance. Anyway you eat toast every weekday morning (you eat pancakes on the weekends), and you're wondering how many shocks you'll get this week. It's a stressful week and you've decided that toast is only worth the risk if you'll probably get shocked only once.

You know the multiplication rule of probability, so you can calculate this. There are five different ways you can receive only one shock this week: you either get shocked once on Monday, Tuesday, Wednesday, Thursday, or Friday and then remain shock free on the 4 other days of the week. If we represent a shock with an X and a non-shock day with an O, the possibilities for your week look something like this.

Now we need to calculate the probability of getting one shock and four non-shocks using the multiplication rule. Let's look at the probability of only getting shocked on Monday. The probability of getting shocked is 20%, so the probability of not getting shocked on Tuesday is 80% and similarly on Wednesday through Friday there's an 80% chance each day of not getting shocked.

So the probability of getting shocked on Monday and not on Tuesday-Friday is 0.2 x 0.8 x 0.8 x 0.8 x 0.8 = ~0.082. That's about an 8.2% chance that you'll get shocked on Monday and not for the rest of the week, so now we've got to calculate the rest of the one-shock options. The probability of getting shocked only on Tuesday is the same, about 8.2%, since order doesn't matter in multiplication.

The probability of the Monday-only option or the Tuesday-only option or any of the remaining 3 options can be calculated using the addition rule. You have 8.2% + 8.2%+ 8.2%+ 8.2%+ 8.2% also known as 5 x 8.2% chance of just one shock. That's a 41% chance of getting shocked only once this work week, so you decide to risk it.

That was a lot of work just to figure out whether it's a good idea to have toast, and thankfully there's a more compact formula. The formula, called the Binomial Distribution formula, takes the math we just did and simplifies it. In our toast example, we first figured out the probability of only getting shocked once.

To do this we multiplied each day's probability together. Let's use exponents to make this formula a little bit shorter. We can combine all the 0.2s and all the 0.8s to give us this.

That's about 8 percent.. Notice that the two exponents add up to 5--the total number of days. This formula works for finding out the chance of getting shocked only once, but we can also use it to find out the chance of getting shocked another number of times.

In general the formula looks like this: For example, the probability of getting shocked only on Tuesday and Wednesday would be this. = ~0.02 or 2% We also need to account for the number of ways that getting shocked once or twice in a week can happen. To do this, we'll use a very useful formula--the Binomial Coefficient Formula--from a field of math called Combinatorics. The Binomial Coefficient formula makes it easy for us to find out how many ways a certain ratio of successes--not getting shocked--to failures--getting shocked, can occur.

In a general form, it looks like this: (n-Choose-k) but you also can read it as "we have n things, count the number of different ways we can choose k of them". For our toast example (5-Choose-1 is 5). That is there are 5 different ways we can only receive one shock from our toaster this week.

The math behind this snazzy formula uses a lot of factorials which are the product of an integer and all the integers less than it, and looks like this: You remember factorials because they're the ones that look like they're always shouting. See those exclamation marks?! We won't dig into this formula here, but for more information, we'll link some resources in the description.

Now we have all the pieces of our binomial distribution formula, let's put them together. First the Binomial Coefficient which tells us how many ways we can have one shock and four non-shocks, and then our shorthand multiplication of probabilities. We put it together and we have a full blown formula for calculating the probability of getting shocked on one out of five days this week, about 40%.

It took us a while to get here, but we now have a general formula for calculating any similar probability. If p is the probability of our event happening then here's our formula: For example, supposing that there is an equal chance of having a boy or girl, if we want to find out the probability of a couple having 3 girls out of 5 children, we can simply plug our numbers into the equation. We do that and we see there's about a 31% chance of having 3 of their 5 children be girls.

And now we get to the zombies: it's the beginning of the zombie apocalypse and you, thankfully, still have your brains. But there's a bunch of people between you and the nearest shelter, and you want to know how likely it is that none of them have been bitten and infected with the zombie virus. Since you can't always tell if a person is infected right away, you decide to use your binomial probabilities to calculate your chances.

It's early stages, so right now there's only about a 5% infection rate in the population. That means the probability of someone being a zombie is about 5%, and the probability of not being infected is still about 95%. Peeking out at the crowd that stands between you and the shelter you count 20 people.

Plugging these numbers into our formula, we can see that there's about a 36% chance none of the people you'd encounter on your run to safety will try to eat your brains. Those are pretty good odds, and you're fast. You know that you could probably safely reach the shelter even if there were one or two zombies, so let's calculate those probabilities as well.

There's about a 36% chance of meeting no zombies, about a 37%

chance of having only one--easily dodgeable-- zombie, and about an 18% chance of having to out maneuver only two, that means that cumulatively, that's about a 91% chance that you'll be able to sprint your way to safety. We could calculate the probability for every possible number of zombies from 0 to 20. If we did that, we would get the discrete distribution for this specific type of problem where there are 20 events, and the probability of the event of interest--here, the zombie infection rate--is 5 %.

But in a more general sense, I also want to know how many zombies I would expect to confront on average - just so I could prepare. I could try to guess the mean by looking at the graph of the binomial distribution for all of the possible cases. Just by eyeballing it, I'd say the mean of this distribution is around 1, and it is!

The actual formula to find the mean of a binomial distribution is n--the number of events we're looking at--times p--the probability of what we're interested in, the probability of being a zombie. Since the probability of being a zombie is 5%, it makes sense that on average, about 5% of any population will be zombies. Since we have a group of 20, we expect about 5% of 20--or 1--zombie infection on average.

While zombies are probably not on your day to day list of concerns, this kind of calculation might be. Important public health issues such as the spread of pandemic-level viruses can been modeled using a similar approach. We could just as easily calculate the probability that of the 40 people you shook hands with at that Zombie Apocalypse meet-up, 2 or fewer had the cold that has been going around.

Or we could calculate the mean number of cold infected people you could expect to shake hands with at any given meeting. If the probability of having a cold at that given moment in the population is 20%, the probability of only 2 or fewer people having a cold is only about 0.7%. From the binomial distribution with 40 people and a 20% chance of having a cold we can see that it's MUCH more likely that more than 2 people will have the sniffles, in fact you'd expect 20% of 40 --or 8 people--to have a cold at any similar meeting with 40 people.

A special case of the Binomial Distribution with only one trial--or just one event, a single coin flip, for example--is called the Bernoulli Distribution. It represents the probability of of getting either a success or a failure. Our outcome--x--can either be a 0 (failure) or a 1 (success).

The general formula for the Bernoulli distribution where p is the probability of success is this. And by plugging in 1 or 0, we can see that the probability that our outcome is a success is p, and the probability of a failure is (1-p). For example, if the probability of rolling an odd prime number (3 or 5) on a single dice roll is about 33.33%, then the bernoulli distribution for rolling a prime number would be the probability of rolling an odd prime to the x power multiplied by the probability of not rolling an odd prime to the one minus x power.

When x is 0--a failure-- then the probability is 1 minus one third or around 66.66%, and when x is 1--a success--the probability is around 33.33%. Though relatively simple, the bernoulli formula is a useful building block. For instance, you can think of the binomial probabilities that we did as many bernoulli trials one after another.

And in reality, that's what they are! Alright let's get back to the zombies. Things have gotten much much worse.

Zombies are everywhere - you're injured really badly. You need a blood infusion and need so much blood that you'll need some from all 3 of the people around you, luckily you're type AB positive so you can get blood from any blood type, but you have something else to worry about: the latent zombie virus. All three of your buddies seem okay, but you heard on the radio the population now has a 30% infection rate.

So there's a 30% chance they're infected with the zombie virus and are just asymptomatic. Luckily, the human immune system can handle the virus in very small doses, so you'll be okay if at least two thirds of the blood is clean. What are the chances you're gonna survive the infusion?

If someone has a latent zombie infection,then the bernoulli trial would have one success and zero failures, giving this formula for the trial: Which as we expect, means that there's a 30% chance that this one random person has a latent zombie infection. And every person who is not infected is considered to be one failure and 0 successes, which has a 70% chance, based on this bernoulli formula: We can use different combinations of these two bernoulli formulas and multiplication to figure out the probability that you'll remain uninfected after your much needed blood transfusion. You'll be fine if no one has any latent zombie virus, and we can calculate the probability of that happening by combining three successful bernoulli trials.

And this looks a lot like the second part of our binomial formula, because it is! A 34.3% chance of survival is kinda low, but don't forget you'll also be okay if only one person has a latent infection, in other words two successes and one failure, which we can combine like this. And there's three ways this could happen (any of your three friends could be infected) so there is a 3 times 14.7--or a 44.1% chance that only one of your friends will be infected.

Combining the 44.1% chance that only one friend is infected, with the probability that all their blood is zombie virus free, we can see that there's a 78.4% chance that this life-saving transfusion will go your way. You're not going to be a zombie… probably. These two formulas are shortcuts to help us spend less time calculating, and more time applying probabilities to our lives.

These concepts also allow us to see how probabilities of a single event can give surprising results when we combine a bunch of them. For example let's say the probability of NOT having a flat tire each year is 95%, meaning there's a 5% chance that you'll get a flat tire. But over 15 years, that's only a 46% chance your tires are going to stay inflated, that means there's a 54% chance of getting a flat tire at some point.

We don't see the world as just one event at a time - many things are connected; So having the tools to think of many probabilities together is important. Thanks for watching. Stay zombie virus free, and I'll see you next time.