Randomness: Crash Course Statistics #17

Crash Course: Statistics
https://youtube.com/watch?v=jL9en6NvQfk
https://nerdfighteria.info/v/jL9en6NvQfk

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

There's a lot of talk of "randomness" in statistics. It's probably something you've heard a lot in this series and in real life too.

Randomness is tied to the idea of uncertainty. Like why are these fries here? And are they delicious?

But just because something is random doesn't mean we know nothing about it. For example, I might not know exactly how many people will shop at my local Costco today, but I do know it's probably more than 100, and probably less than 1 million. Even with these very conservative guesses, I still know something about the "randomness" of this variable.

It's an odd juxtaposition of what we know and what we don't at the same time. INTRO Lots of things in your everyday life are random. From dice rolls in your weekly Dungeons and Dragons game, to which card you draw next when playing Canasta, to how many people in your subway car are Trekkies.

Since individual values of these random variables aren't that predictable, we generally look at the outcomes across multiple instances. Often, the best way to get a feel for the behavior of random variables like dice values is to simulate them. Simulations allow us to explore options that didn't happen, but could have happened.

And when you get right down to it, that's what statistics is all about. Let's look at a simulation to understand more about the weight of a Large Fry at McDonald's. Supposedly, a Large Fry at McDonald's has about 168 grams of crispy, salty, potato-y goodness.

But we know that the process of shoveling piping hot fries into cardboard cartons, isn't an exact science. But it most likely is a random process, which means that the weight of your fries, is a random variable. You don't know exactly how many grams will be in your next order of fries, but you can know something about the random process that generates these weights, so you hop in your car, drive to the nearest McDonald's and order 10,000, large fries.

Back at home you get out your scale and begin to unbag and weigh your fries… the first batch weighs 173.03 grams. Not bad, seems like you got an extra fry or two. The next 4 orders weigh 169.05 152.41 153.80 174.60 grams respectively.

After unbagging and weighing all 10,000 orders you plot a histogram of all the weights. Looking at this graph, we can see that McDonald's is pretty good at giving you your fair share of fries. Most orders have around 168 grams.

But we can also see that the randomness of the carton-filling process means that we can expect to occasionally see orders with up to 200 grams and as low as 130 grams... but those don't happen too often. You may ask how many grams of fries you should expect to get on your next trip to McDonald's... our best guess is the mean. It's the amount we expect to get from this random fry distributing process.

You already know how to calculate the mean of a finite group of numbers--the sum of those numbers divided by the numbers of items, n--so let's expand our definition so that we can take the mean of a distribution. Again, the mean is a type of expectation, it's called the expected value of the data because it's what we "expect" from our data overall. Like how you "expect" that an American woman would be the average height of 5'4 or 163 centimeters if you didn't know anything else about her.

With discrete distributions, where values can only take on set numbers like how many sodas people drink at a party... calculating the expectation looks similar to the mean formula. For simplicity's sake, let's say that people will only take 0,1,2, or 3 sodas at your party. And you want to know how much soda you should expect a person will drink so that you can get enough.

Cause nothing kills a party like running out of soda… For each possible value, multiply it by the relative frequency for that value and add all these products together. So if 10% of people will drink 0 sodas, 20% will drink 1 soda, 40% will drink 2, and 30% will drink 3 sodas, we get this formula for the expected value. Which equals 1.9 sodas, meaning you should buy about 2 sodas per guest.

Notice that we didn't have any actual counts for your guests...no one RSVPs anymore. This is the expected value of the distribution, and we can apply it to any number of guests we want. But not everything in life is measured discretely, sometimes...oftentimes… you'll have continuous variables like height, or grams of fries, which can take on any value at all.

In theory, calculating the expectation for a continuous distribution is exactly the same, except now we have an infinite number of values which means adding all of the products of values and frequencies isn't really doable. Luckily Sir Isaac Newton invented the integral which allows us to take the sum of an infinite number of these products without actually adding them all up by hand. You may see it written like this.

But this is simply the fancy math way of saying, "multiply all the values by their frequencies and add 'em up". If we wanted to know the expectation of the weight of a large fry in grams, we can use this integral and the fact that the fry weights are normally distributed to calculate it. No matter how you're calculating it, the expectation of your data is an important thing to know.

It not only characterizes the data, but it can help you make sure that you know what to expect, from number of sodas to have at a party, to how much joy you should expect from a belly full of fries; expectation helps you understand something about randomness. But not everything...there's still more to know about random processes, like how spread out or how variable they are. Variance is also an expectation.

It tells us how spread out we expect the data to be. The variance of the amount of money each family makes is pretty high, because people don't all make the same amount of money. To make things easier, we can represent expectation like this.

Variance is the expectation of all data points minus the mean squared. Since we're subtracting the mean, we call it mean-centered or "central". In essence, we're creating a new distribution (each value minus the mean squared) , and taking the expectation of this new distribution.

Since expectation is always the same--we just sum a bunch of values times their frequencies -- these two formulas are the same. Since we're taking each value minus the mean to the second power, we also often call this the second moment of the data. Which is just the expectation of the mean-centered data to the second power.

The second moment tells us how reliable the first expectation is...if you have a really high variance for your estimate of how much soda to buy for your party, you know that you might want to run to the store for a couple extra cases, since it's possible that you might get a group of real soda guzzlers. So the mean is the first moment of a distribution of data, and the variance is the second. And we can keep going.

There are a lot of moments since all we do is keep raising to higher and higher powers, but the first four are the most useful for our purposes. We've already covered the first two, but the third moment--the expectation of the mean centered data to the third power--is also something you might be familiar with: Skewness. Skewness tells us whether there are more extreme values on one side, like income or amount won in Vegas which are both right skewed.

Think back to your algebra class...when you take something to an even power like 2, or 4, your number is always positive. So even moments--like variance--are always positive. Variance counts extreme values on the right and the left of the mean the same, since it squares them. -2 squared and 2 squared are both 4, so values that are 2 units above or two units below the mean both contribute equally to the variance.

But odd powers like 3 can be negative or positive, so they count numbers above the mean differently than those below. Numbers smaller than the mean that are negative will still be negative when they're taken to the third power. So the third moment--skewness--is a measure of how skewed the distribution is.

If there are a lot more extreme values smaller than the mean, skewness will tend to be negative. On the other hand, if there are a lot more extreme values bigger than the mean, skewness will tend to be positive. We've seen that as humans, we're pretty good at seeing when a distribution is skewed, but it can be really useful to have a way to quantify it.

Just like the variance tells us how reliable the mean is, skewness can tell us how reliable the variance is. If a distribution is really skewed, then the variance is going to be a lot higher on one side. Imagine the distribution of the amount of chips that people will eat at your party is skewed, you know that there's a lot more extreme values on one side...maybe some people forgot to eat dinner before they showed up.

And finally, the Fourth moment - Kurtosis. Kurtosis is the Expectation of the mean centered data to the fourth power. And it's a measure of how thick the tails on a distribution are.

This tells you how common it is to have values that are really far from the mean. When you're playing music at your party, the distribution of how loud people like the music to be might have high kurtosis; There's a lot of people who want it quiet so they can talk, and others who ...don't want to talk. Though it's not as common, kurtosis, along with all the other moments, can help us have more information about a random distribution.

For example, it can help us tell whether a variable follows a normal distribution. You can see the mean--the first moment--tells us where a distribution is on a number line. When you change the mean, you slide the distribution left or right.

The other moments tell us about the shape and spread of the distribution, which stay the same no matter where we move the distribution. So it might make sense that when we add two independent random variables together, like the sum of two dice rolls, the mean of this new distribution is the sum of the means of the two distributions being added. And this is true no matter how many means you add.

Maybe your stats teacher has said "The mean of the sum is the sum of the means". Similarly the variance of the sum of two independent variables is the sum of their variances. So if we do want to look at the distribution of the values of two dice rolls, we can easily calculate the mean and standard deviation.

The mean of one die roll is (1+2+3+4+5+6)/6 or 3.5 The mean of two dice rolls would be 7, since it's the mean of the first roll, plus the mean of the second roll. The variance of the value of one roll is about 2.9 which means the variance of the value after rolling two dice is about 5.8. And as for those fries, we'd expect to get about 336 grams if we ordered two larges.

Randomness is the reason you can't be sure you'll win in Las Vegas, or why you always have to leave early because you can't predict how long you'll have to wait for a parking spot, or why sometimes you bring an umbrella with you on days when it doesn't end up raining. But the beauty of statistics is that it helps us know something about this randomness and make better, more informed choices in the midst of chaotic randomness. Like deciding whether a machine learning algorithm trained to recognize sheep is truly better than humans at recognizing sheep in unusual places.

Or even whether the increase you observed in fecal matter on people's hands is really higher after using air dryers than paper towels. Thanks for watching, I'll see you next time.