Z-Scores and Percentiles: Crash Course Statistics #18

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

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

One thing that statistics are good for comparing. You can compare your GPA with the mean or median GPA, and you can use the standard deviation to figure out whether the amount of time that people spend on social media everyday is pretty similar, or whether people differ a lot.

But both these examples are comparing apples to apples. And sometimes we want to compare things that aren't exactly the same, or aren't measured in the same way. This is where standardization comes in.

For example some prospective college students took the SAT, some the ACT and others both. How can we compare things that aren't the same? It's like comparing apples to grapefruit.

INTRO Say there are two students who are applying for admission to the same college-- one took only the SAT, and the other only the ACT. Both tests are trying to measure the same thing: college readiness. But they're different tests, and more importantly, are measured on different scales; The SAT is currently out of 1600 points, while the ACT is out of 36 points.

This makes things tough to compare. While I can maybe assume that a perfect score on the SAT and ACT mean similar things--namely that I've found a superstar test taker--it's not immediately clear whether a 1200 on the SAT is better than a 25 on the ACT. Again, apples to grapefruit.

They're similar but not quite the same. The first thing we can do To make these scores easier to compare is to center both of the distributions around zero by subtracting the mean of each respective test from each score. Now for both adjusted test scores, someone who got around the mean of either test will have a score of 0.

So, if you scored a 1000 on the SAT or a 21 on the ACT , your new, adjusted score would be 0, since those scores are 0 points away from the means of each test. Now that both scores are centered around zero, it's a little easier to compare them. Things that are close to zero indicate a score that was close to the mean, while those far away from zero indicate scores that were either a lot higher, or a lot lower than the mean score for that test.

But our test scores are still on different scales. A 10 (above the mean) on the SAT obviously isn't the same as at 10 (above the mean) on the ACT. The second, and final step that will allow us to compare these two scores is to take our already adjusted scores and measure the distance away from the mean using units of standard deviation.

We do this by dividing the adjusted scores by the standard deviation of the respective tests. This rescales both distributions of scores so that the standard deviation for both is now 1 unit. A re-scaled score, which is called a z-score, of 1 indicates a point that is 1 standard deviation higher than the mean, and a z-score of -1 indicates a point that is 1 standard deviation lower than the mean.

Now we can better compare two students: Tony who got a 1200 on the SAT, and Maia who got a 25 on the ACT. First we subtract the mean score from each, and then divide by the standard deviation. Let's say those standard deviations are 200 and 4.8 for the SAT and ACT respectively.

So Tony's score of 1200 minus 1000 divided by 200 gives a new z-score of 1 (which means means that it is 1 standard deviations above the mean). And Maia's 25 becomes 25 minus 21 divided by 4.8 giving a new z-score of 0 .83 (which means that her score is .83 standard deviations above the mean). These scores are much more comparable.

It's easy to see that Tony and Maia's z-scores are actually pretty similar, even though it was hard to tell with their original scores of 1200 and 25. Z-scores in general allow us to compare things that are not on the same scale, as long as they're normally distributed. In fact there's an entire z-distribution that allows us to calculate things like percentiles.

Percentiles tell you what percentage of the population has a score or value that's lower than yours. The median, is the 50th percentile; exactly half the data is above and below it. By looking up Tony and Maia's z-scores in a standardized z-table, we can see that Both scored above approximate 80% of their peers give or take.

You may have seen percentiles at the doctor's office when your physician told your parents that you were at the 83rd percentile for height. That means that 83% of the population-- and here we mean kids your age, not all people ever--have heights that are less than or equal to yours. If you look at the whole distribution of heights and found yours, 83% of the distribution would be to the left of your height.

Now, technically you can find these percentiles for any distribution, it involves calculating the area underneath the curve and figuring out how much of it is below a certain value. But that calculation can be tedious, so often times we convert normal distributions into standard normal--or z--distributions which have a mean of 0 and a standard deviation of 1. Common percentiles have already been calculated for the z-distribution, you may have seen a table of them in a statistics textbook.

Say you are competing for a spot at a local video game convention. The rules state that you must be in the top 5 percent of scores for your favorite game, Call of Civic Duty--a jury duty based video game. Scores of Call of Civic Duty are normally distributed with a mean of 2,000 and a standard deviation of 300.

You get 100 points each day you don't fall asleep. So you set out to find out what score would put you at the 95th percentile. Being in the 95th percentile means that you're in the top 5 percent of scores since 95% are below you.

And these are just two different ways of talking about the same situation. Looking at our z-score table, we can see that a z-score of about 1.65 would put you in the top 5 percent of Call of Civic Duty players. In fact a z-score of 1.65 corresponds to the 95th percentile of any z-scored distribution.

But we want to turn that back into a Call of Civic Duty score so that you know just how high you have to score in order to qualify for the gaming convention. We know the process to turn a raw score into a z-score, so to go the other direction, we just reverse it. First, we multiply the z-score by the standard deviation--300--to get 495.

Then we add the mean score 2,000 to get the final score to beat: 2,495. We can also think of percentiles as the probability that a random score drawn from a z-distribution will be lower than a given score. Say, you are at the 90th percentile for marathon runners.

That means that if we randomly selected a marathon runner, 90% of the time we'd select a runner with the same or slower run times than you. So you're in the top 10 percent, but does that mean you're "extremely" good? There's really no clear answer of how good you have to be or how high a score has to be in order to be considered "extreme", it's somewhat arbitrary.

You may think the top 10% of marathon runners are "extreme", but

someone might draw the line at the top 1%. In fact, sometimes you can be so extreme that people might start to think you're not even from a certain population z-distribution. Let's go to the Thought Bubble.

You're at the county fair, and stop at a game booth. The woman running it says that she has two piles: one of apples, and one of some type of mystery object. She'll randomly pick an item from one pile.

If you can guess whether she picked an apple or not, you'll win a lifetime supply of funnel cake. She does kindly tell you that the mean weight of this type of apple is 200 grams with a standard deviation of 20 grams. She grabs an object, weighs it, and yells out the weight of 270 grams.

You quickly calculate the z-score is 3.5. You take out your phone and pull up a z-score table and figure out this object would be in the top 99.99th percentile, meaning that it is larger than 99.99 percent of apples. It's such a big apple, it's basically New York.

If it was an apple, it's unlikely you'd get one this big, only 0.001% of the time. So maybe it's not an apple. Before you give your final answer you need to ask yourself whether you think that's rare enough for you to conclude that it's probably not an apple, the limit is pretty arbitrary, but it's enough here for you to guess not an apple.

And you swear you see a twinkle in the proprietor's eye when she tells you that you're wrong and pulls out a very large apple. As you walk away, you think about how nice it would have been to know what the mystery pile was. If it had been a pile of Bowling Balls, you probably would have guessed "apple", even though it seemed a little heavy since it's more likely to have a 270 gram apple than a 270 gram bowling ball.

But if the other pile were grapefruit, you might be more inclined to stick with your original choice of "not an apple", since a 270 gram grapefruit is more likely than a 270 gram apple. But since you didn't know, you made the best choice that you could and you can buy your own funnel cake. Thanks Thought Bubble.

Z-scores help us make comparisons. We can compare a thing to a certain population, like a particular apple weight to all apple weights. And we can compare stuff that's not naturally comparable like SAT and ACT scores.

Though we might not calculate actual scores, the ideas behind z-scores are what make us feel that the likelihood of being on the same flight as Chadwick Boseman is way smaller than the likelihood of hitting 4 red lights on our way home, even though those things are very different. Or maybe we're trying to compare athletes from different sports to see who the GOAT is, no not that kind of goat, Greatest Of All Time. Let's say our top contenders are Lebron James and Tom Brady.

But who's the GOAT? Lebron has a career average of about 27 points per game, while Brady has an average of 1.92 touchdowns per game between 2000-2017. While both are impressive, we're looking for the greatest.

When we're picking out our GOAT, we could compare the two athletes by thinking about how much higher above the average score each one has. But to know which athlete's score is more impressive we'd need to compare their z-scores. I haven't run the math, but off the top of my head, I'm going with LeBron.

Prove me wrong. Best analysis wins something that weighs less than 270 grams. No other clues.

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