Robert_S2_L08

Wed, 1/12 12:25PM 14:19

SUMMARY KEYWORDS

quartile range, median, range, values, observations, inter, divide, frequencies, inches, iqr, children, calculating, number, family, variability, highest paid player, 25th, 75th percentile, basketball players, information

SPEAKERS

Robert McKeown

R Robert McKeown 00:07

Welcome back. Here's an example that teaches us to be careful when it comes to finding the range. We've got a bar graph, and we've got number of children on the X axis. So we've got a family here with no children, and a family here with eight children, and all numbers of children in between. We have the values, the frequency, the frequent occurrence of each of these family sizes, stated up here at the top of the bar, and the question just tells us to find the range. And so you might think, Oh, well, the range is eight minus zero, right? There's nine different possibilities for family sizes. And so the range is just going to be eight, because 00, so eight minus zero, gives leaves us with eight. But that's not exactly the way that the question probably wants us to interpret itself. What the question is probably asking us is the range in terms of frequencies. So the range might not actually be in terms of the number of children, although that seems very intuitive. The range in terms of the frequency of different observations could be something like, looking at the frequencies. So we've got 14 here, and we've got 391. There. And so the range in terms of the frequencies of the each of these occurrences is going to be 391 minus 14, which is going to give us 377. Two kind of a, it's a trick question. In some sense, the number of children and almost think of it as categorical data. It has to be an integer and no family in this sample had more than eight children. But we still need to be so we want to be careful when we're calculating range, what exactly what range? Are we really looking at here is something to consider when we're asked these questions. Now, the range has its problems, it's simple. And it only provides information about the endpoints. And those endpoints might be extremely large or very small. So when we saw the Toronto, Raptors basketball players, we saw that, you know, one player was making $33 million. And then the next highest play paid player was making $20 million. And so you would have a very large range, but you wouldn't really know the distance between, say, the highest paid player and the second highest paid player. When we looked at the Toronto Maple Leafs, the highest paid player had 11 point 6 million, but the second highest paid player had 11 million. And so these things might be of interest to us. But the range doesn't really tell us anything about the values between these two endpoints. And so the range provides no information on the interior values. The standard deviation and variance are more common measures of the spread, or dispersion variability. And we'll look at those in a later video. Now there's something called the inter quartile range, it seems to be very popular in business. And what it does is really focus on

the middle of the the distribution or the middle values. Like so you could divide it into four quarters. So you might say, Oh, this is quarter 1, 2, 3, 4. And so sometimes the year, for example, is divided into four quarters. And each quarter is three months long as the inter quartile range tells us what the distance is from here to here, so it's going to be from the the middle 50% It's that middle 50%. We're looking at that middle 50% And we want to find The value at the 75th percentile. And we want to have the value at the 25th percentile. And the IQR is going to be equal to the value at the 75th percentile minus the value at the 25th percentile. So the inter quartile range divides the set of values into the four quarters. One quarter of the values in the set are below the lower quartile, and one quarter of the values in the set are above the upper quartile. So if we go back to look at this diagram, we've got 25% over here, which is outside the IQR. And we've got it 25%. Above that 75th percentile, which is also going to be outside the inter quartile range there are a few ways to find the inter quartile range. So we'll work through this example together is to start by finding the median. And so we've got some data here. And we're going to work on this chart to illustrate, we've got soccer player heights, and they're listed from lowest, to highest. And they're stated in inches. Our first step is to find the median. So we're going to find median. And to do that, we need to know how many observations there are, well turns out there are 101 observations. And the frequencies are on the diagram as well. So we have the frequencies at the top of the bars. And the median here is going to be the number of observations plus one divided by two, which is going to give us 51. So the 51st observation is going to have 50 observations below and 50 observations above. Where is the 51st observation while we have to do a fair bit of counting. And we can see we've got six observations here. We've got 15, 3, 18, here, we're getting closer, and then we've got 10 there, so we've got 24, 24, 34, 45. And then if we added these 12, we would end up with 57. So it's going to be somewhere in this 12, which we count all the way down looks like it's going to be between 67 and 67.5 inches. So median is going to be somewhere, let's just say 67 inches, simplicity. Now step two is find the median the lower. So we want to look at all the observations over here, basically, and we want to find the median of that area. Now we know there are 50 observations. And if we want to find the median, we're going to take 50 plus one and divide by two, which is going to give us of course, we have an uneven number of observations. So we're going to have two median values, and they're going to be the the 25th and the 26. observation. So if we take these out on a limb, and we've got six here, nine, plus 716 plus 824. It looks like our 25th and 26. Observations are both going to be 66 inches, so we've got a lower median. 66 inches and step three is to do the same thing. To find the median the upper half now it's getting a little tricky. So what I'm going to do is I'm going to, you know, we want to find the median of this area in here. And I'm going to start counting from the right to left until I get to the 25th and 26 observations, so about 2, 1, 5, 8, then we've got 11, 14, 20, 25, and 26. So it looks to me, like the upper half median is going to be 69.5 and 70 inches. Remember, if we have an even number of observations, in this case, we had 50. That means we can have two median values. And if I draw a dotted line, up here, we can see that our inner quartile range is going to lie. Oops, it's gonna lie. All in here. So we're going to kind of color this in. Like so. And that's our this is our inter quartile range here, calculating the value, we're going to get 70 minus 66, which equals four inches. So it's going to be equal to four inches, or if we use the second median. 3.5 inches, and that is our inter quartile range. Not common that you would calculate this by hand. Usually, you would let a software package calculated for you. As you can see, it's kind of awkward the calculations to do it yourself, or rather awkward. Here are the steps written out. There are the soccer player heights, you can color it in yourself and identical and identify the IQR. Why would you want to use the inter quartile range? Well, it gives you more information than just the range, or at least it gives you information about what's happening at the middle of the distribution. And that could be particularly important if you have extreme values, also known as outliers. And that might be influencing the range to make it seem bigger than the actual variability of the series, it might be useful to compare the

range and the inter quartile range to see if you can identify whether there might be extreme values at the lower end or the upper end that are influencing the range and making it seem like the series is more varying than it actually is. Occasionally, researchers are entered interested in quintiles so they'll separate this the the series into five parts sections instead of fours. And sometimes they'll divide the series into deciles. So that divided into 10 parts. These are going to be a little bit more complicated to measure but the idea is the same, but software packages are going to be able to help you do this