

The student is working on the following problem:

Table 1. shows the number of goals that each football player striker scored between 2007 and 2011. Who is the most consistent player? Immagine a measure of "consistency" and justify why you think your method is correct.

Year	Mike Arwen	Dave Backhand	Ivan Right
2007	13	12	14
2008	12	14	10
2009	15	16	18
2010	17	15	18
2011	13	13	15

The solution of this problem is:

The concept of standard deviation is the usual way to measure consistency but it is not expected that the student knows it. Any measure proposed by the student to measure consistency is acceptable as long as it can be justified.

Example of canonical solution: computing the variance (or standard deviation) for each player: First, compute the mean:

Mean number of goals for Mike: 14

Mean number of goals for Dave: 14

Mean number of goals for Ivan: 15

Then, compute the sum of square deviations from the mean for each player.

Then devide by the number of data points to get the variance:

Variance for Mike: $12/5 = 3.2$

Variance for Dave: $10/5 = 2$

Variance for Ivan: $44/5 = 8.8$

So according to the variance, Dave is the most consistent player.



Hello! I am your tutor. Can you walk me through your solution?

Oh hi! I'm actually not sure how to start. The problem says to measure consistency, and I don't really know what that means in terms of math. What's a good way to measure something like that?



I understand! Think of "consistency" in everyday terms first—what does it mean for someone to be consistent in their performance over time? When you have some idea about that, try to express those thoughts in a mathematical way. What could indicate a consistent performance when you look at numbers?

Hmm... I guess consistency would mean that someone is doing pretty much the same thing over and over, without a lot of change. So maybe for the player scores, it would mean their number of goals doesn't vary much from year to year. The closer the goal numbers are to each other, the more consistent the player is. How do we turn that into math?



You're on the right track! Now consider how you might quantify the difference between the numbers for each player. How could you express the variation in their yearly goal counts mathematically? Think of simple calculations you could perform on the numbers to represent the idea of "closeness" or "sameness".

Ok, so if the numbers were all the same, there would be zero difference between them. But since they're not, we could find the differences between each year's goals and some kind of average goal number. Maybe by subtracting each year's goals from the average? And then we could take those differences and do something with them to get a single number that tells us how much the scores vary. Does that make sense? What should we do with the differences?



You've pretty much described the first steps in a very important statistical process. Now, you have these differences from the average for each year. They might be positive or negative, depending on whether the actual goals are above or below the average. What could you do with these differences to deal with the fact that they can be both positive and negative, and yet come up with a measure that reflects the variation in a clear way?

Oh, I got it! If you square the differences, they all become positive. So you wouldn't have to worry about positive or negative anymore because squared numbers are always positive. After squaring the differences, maybe add them all up to get a total measure of how much it varies?



Brilliant! You've squared the differences and now have positive values. You have a sum that represents the total variation from the average. Now think about how you can make this sum representative of all the years, not just an absolute total. What can you do with this sum to achieve that?

Since we squared and then summed up the differences for each year, we could divide that total by the number of years to get like an average of those squared differences. So it would be like sum of squared differences divided by... the number of years is 5, right? Because the data is for 5 years, so we divide by 5.



Exactly, you're dividing by the number of years to find an average. What would this represent in the context of the players' performance? And what would your next step be to complete this measure of consistency?

So if I do that, I'm getting an average of those squared differences for each player, which should say how much their goals typically deviate from their average goals over the 5 years, in a squared way though. But those numbers would be kind of big and weird because they're squared. Shouldn't I take the square root of that average so it goes back to being sorta like the original numbers? Would that make it better?



You're considering an important point about the scale of the numbers. If you take the square root of your average of squared differences, what kind of value would you obtain and why might that be a useful thing to do? Think about how it connects back to the original goal counts.

Taking the square root would give me a value that is more like the original goal numbers because it wouldn't be squared anymore. It's like undoing the squaring I did before. That way, the value is sort of like an "average" difference from the average goal count, which sounds like it tells you how much a player's goals are spread out from their average. So the lower that number, the more consistent they are because they don't spread out as much from their average. Is that called something? Am I supposed to give it a name?



What you've cleverly derived is indeed a well-known statistical measure. Considering the process you've followed and the terms you've used ("average" and "difference"), can you think of a statistical term that might describe this value you're proposing? How would you describe it in the context of assessing the players' consistency?

Oh, I think I've heard of something like that. Is it the standard deviation? So that would be like the standard amount that a player's goals deviate from his average each year. For consistency, a lower standard deviation would mean the player's goals are closer to his average all the time, so that could be a measure of consistency. The player with the smallest standard deviation would be the most consistent, right?



Correct, you've got the concept of standard deviation. Now, with that understanding, how would you apply this to the players' data to complete your measure of consistency?

Alright, I think I need to calculate the standard deviation for each of the players.



Here's what I should do:

1. Calculate the average goals for each player over the 5 years.
2. For each year, subtract the player's average goals from their actual goals to find the difference.
3. Square each of those differences.
4. Add up all the squared differences for each player.
5. Divide the sum by the number of years to get the mean of the squared differences (this is the variance).
6. Take the square root of the variance to get the standard deviation.

Should I go ahead and start these calculations with the goal data?



It sounds like you have a clear plan regarding what steps to take next. How do you feel about proceeding with these calculations using the players' goal data? What do you think would be your next step to move forward?

I feel good about it! Let's do the steps for each player.



First step is to calculate the average goals for each player. Let's start with Mike Arwen. Mike's goals are 13, 12, 15, 17, and 13 for the five years.

So Mike's average goals would be:

$$\text{Mike's Average Goals} = \frac{13 + 12 + 15 + 17 + 13}{5}$$

$$\text{Mike's Average Goals} = \frac{70}{5}$$

$$\text{Mike's Average Goals} = 14$$

Do you want me to calculate the average for Dave and Ivan too? Or should I continue with the next steps for Mike first?



Great job with Mike's average! What do you think would be the best approach for ensuring consistency in your calculations? Should you focus on each player one at a time or calculate the same step for all players before moving on?

I think it might be easier to calculate the same step for all players before moving on. So, let's calculate the averages for Dave and Ivan too. Then after that, we can subtract, square, and do the rest for each one.



Alright, for Dave Backhand the goals are 12, 14, 16, 15, and 13.

So Dave's average goals would be:

$$\text{Dave's Average Goals} = \frac{12 + 14 + 16 + 15 + 13}{5}$$

$$\text{Dave's Average Goals} = \frac{70}{5}$$

$$\text{Dave's Average Goals} = 14$$

And for Ivan Right, the goals are 14, 10, 18, 18, and 15.

So Ivan's average goals would be:

$$\text{Ivan's Average Goals} = \frac{14 + 10 + 18 + 18 + 15}{5}$$

$$\text{Ivan's Average Goals} = \frac{75}{5}$$

$$\text{Ivan's Average Goals} = 15$$

Now I've got the average goals for all three players. We can move on to calculating the differences and squaring them. Should I do that next?



You've calculated the averages with precision. Considering the steps you mentioned earlier, what do you think your next step should be after computing the averages for each player?

Okay, now I'll subtract each player's average goals from their actual goals for each year, square each of these differences, and then add those up for each player.



Let's do that next for Mike, Dave, and Ivan.

Starting with Mike, his average was 14 goals. His actual goals were 13, 12, 15, 17, and 13.

So the squared differences for Mike would be:

$$(13 - 14)^2 = (-1)^2 = 1$$

$$(12 - 14)^2 = (-2)^2 = 4$$

$$(15 - 14)^2 = 1^2 = 1$$

$$(17 - 14)^2 = 3^2 = 9$$

$$(13 - 14)^2 = (-1)^2 = 1$$

Adding those up:

$$1 + 4 + 1 + 9 + 1 = 16$$

Then, we divide by the number of years (5):

$$\frac{16}{5} = 3.2$$

And take the square root:

$$\sqrt{3.2} \approx 1.78885$$

So the standard deviation for Mike is approximately 1.78885.

Should I do the same now for Dave and Ivan?



How do you feel about the calculations you've made for Mike? What is your confidence level on repeating the same process for Dave and Ivan?

I feel pretty good! I think I've got the hang of it now, so I should be able to repeat the process for Dave and Ivan.



Let's go ahead with Dave next, whose average was also 14 goals. Dave's actual goals were 12, 14, 16, 15, and 13.

So for Dave, the squared differences are:

$$(12 - 14)^2 = (-2)^2 = 4$$

$$(14 - 14)^2 = 0^2 = 0$$

$$(16 - 14)^2 = 2^2 = 4$$

$$(15 - 14)^2 = 1^2 = 1$$

$$(13 - 14)^2 = (-1)^2 = 1$$

Adding those up:

$$4 + 0 + 4 + 1 + 1 = 10$$

Then, we divide by the number of years (5):

$$\frac{10}{5} = 2$$

And take the square root:

$$\sqrt{2} \approx 1.41421$$

So the standard deviation for Dave is approximately 1.41421.

Now let's do it for Ivan, whose average was 15 goals. Ivan's actual goals were 14, 10, 18, 18, and 15.

The squared differences for Ivan are:

$$(14 - 15)^2 = (-1)^2 = 1$$

$$(10 - 15)^2 = (-5)^2 = 25$$

$$(18 - 15)^2 = 3^2 = 9$$

$$(18 - 15)^2 = 3^2 = 9$$

$$(15 - 15)^2 = 0^2 = 0$$

Adding those up:

$$1 + 25 + 9 + 9 + 0 = 44$$

Then, we divide by the number of years (5):

$$\frac{44}{5} = 8.8$$

^

And take the square root:

$$\sqrt{8.8} \approx 2.96648$$

So the standard deviation for Ivan is approximately 2.96648.

Based on the standard deviations, Dave is the most consistent player, followed by Mike, and then Ivan.



Reflecting on your process, how does the standard deviation help us understand the goal-scoring consistency of the players? How does it feel to compare the values of standard deviation you've calculated for each player?

filename

ok/BX1

Load Chat History