

What is Standard Deviation?



Teacher Notes and Answers

7 8 9 10 11 12



TI-Nspire



Investigation



Student



30 min

Introduction

Jamal and Rudi play for a local cricket club. Their first four lots of scores for the season are:

Jamal: 29, 35, 29 and 27

Rudi: 7, 12, 18, and 83

This exploration compares the scores for each cricketer and looks at the ways their scores are similar and the ways in which they are different. It also explores the usefulness of a new statistical measure, the *standard deviation*.

Part 1: Exploring deviation

Question 1.

Copy and complete the following to calculate the mean score for each boy.

a) Jamal's mean batting score = $\frac{29 + 35 + 29 + 27}{4} = \frac{120}{4} = 30$

Rudi's mean batting score = $\frac{7 + 12 + 18 + 83}{4} = \frac{120}{4} = 30$

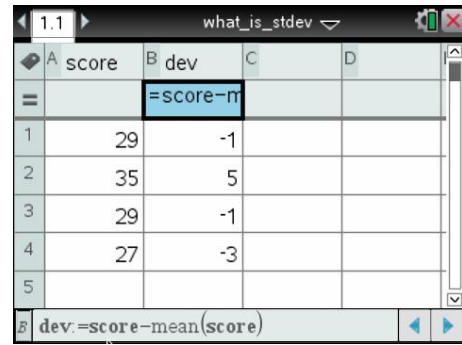
- b) What do the mean scores tell us about the boys' scores? What do the mean scores not tell us about the boys' scores?

Both have scored the same total of runs, and the same number of innings. However, it does not tell you how consistent the batsmen were during the four innings.

The mean provides a measure of the centre of a statistical variable, but not information about how spread out the scores may be. To measure the 'spread' of the data, we will look at how much each individual score deviates (is different) from the mean score.

On the TI-Nspire CAS:

- Press HOME-1 to create a New Document, and then press the relevant icon to add a **Lists & Spreadsheet** page.
- Type **score** as the variable name for column A, then enter Jamal’s scores into this column.
- Type **dev** as the variable name for column B, then in the cell below, type in the formula **=score-mean(score)**.



	A score	B dev	C	D
=		=score-m		
1	29	-1		
2	35	5		
3	29	-1		
4	27	-3		
5				
Ⓜ	dev = score - mean(score)			

Now use your spreadsheet to answer the following questions.

Question 2.

- Which score deviates the most from the mean score?
35
- Which score is closest to the mean score?
29

Question 3.

- A possible measure of the ‘average’ deviation is the ‘mean’ deviation, which is calculated as follows (copy and complete).

$$\text{Jamal's mean deviation} = \frac{\text{Sum of deviations}}{\text{Number of scores}} = \frac{-1 + 5 + -1 + -3}{4} = \frac{0}{4} = 0$$

- Is this a useful measure of the ‘average’ deviation? Why/why not?

The positive and negative deviations cancel out. The mean score will always be the value for which this cancelling occurs. Hence the mean deviation is not a useful measure of the average amount of deviation.

Another possible way of measuring the ‘average’ deviation is to square each deviation first, find the mean of the squared deviations, and to find the square root of this.

On the TI-Nspire CAS

- Type **sqdev** as the variable name for column C
- In the cell below, type in the formula **=dev²**.

Note that all squared deviations are positive.



	A score	B dev	C sqdev	D
=		=score-m	=dev^2	
1	29	-1	1	
2	35	5	25	
3	29	-1	1	
4	27	-3	9	
5				
Ⓜ	sqdev = dev ²			

The mean ‘squared’ deviation of Jamal’s scores can be calculated in the following manner.

Question 4.

Copy and complete the following, giving your answer in simplest fraction form.

$$\text{Jamal's mean squared deviation} = \frac{\text{Sum of squared deviations}}{\text{Number of scores}} = \frac{1 + 25 + 1 + 9}{4} = \frac{36}{4} = 9$$

To calculate the mean ‘squared’ deviation of Jamal’s scores, do the following. on the TI-Nspire CAS

- Move to cell D1, and type the label **meansqdev**.
- Move to cell D2, and type the formula **=approx(mean(sqdev))**.

Confirm that it gives the same answer for the mean ‘squared’ deviation as the one you found above.



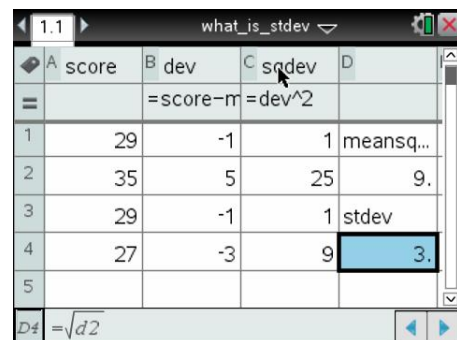
A	score	B	dev	C	sqdev	D
=		=score-m		=dev^2		
1	29	-1		1		meansq...
2	35	5		25		9.
3	29	-1		1		
4	27	-3		9		
5						
D2	=approx(mean(sqdev))					

Finally, to find the ‘average’ deviation by this method, find the square root of the value of the average squared deviation.

Now we have a standardised measure of the mean deviation (referred to as the standard deviation), which ignores whether the individual deviations are positive or negative.

To calculate the standard deviation of Jamal’s scores on the TI-Nspire CAS

- Move to cell D3, and type the label **stdev**.
- Move to cell D4, and type the formula **=sqrt(D2)**



A	score	B	dev	C	sqdev	D
=		=score-m		=dev^2		
1	29	-1		1		meansq...
2	35	5		25		9.
3	29	-1		1		stdev
4	27	-3		9		3.
5						
D4	=sqrt(d2)					

Now lets look at the spread of Rudi’s batting scores (7, 12, 18 and 83 runs).

On the spreadsheet, change the values in Column A to Rudi’s scores.

Question 5.

For these new scores, find (correct to 2 decimal places)

- a) the mean squared deviation

951.5

- b) the standard deviation

30.8

Question 6.

Compare the standard deviation of Rudi’s scores with the standard deviation of Jamal’s scores found previously. What do you notice? What does this suggest?

The standard deviation of Rudi’s scores is much greater, indicating that his scores are, on average, further from the mean score.

Now, suppose Rudi is about to bat again for the fifth time. Try different scores for Rudi's fifth score.

Question 7.

- a) What fifth innings score will make his standard deviation as low as possible?
30
- b) What is the value of this standard deviation (correct to 2 decimal places)?
27.59
- c) Explain why this score will achieve the smallest possible value for the standard deviation.
The numerator in the expression for mean squared deviation is unchanged but the denominator decreases by one. Any other change would increase the numerator also, so the standard deviation would be larger.

Part 2: Exploring the standard deviation

We now look at some other data to explore the idea of standard deviation, and how it is affected by the spread of the data values.

Using the TI-Nspire CAS spreadsheet template you have developed, answer the following questions.

Question 8.

- a) Find 4 batting scores that have a mean score of 30 and a standard deviation of 0.
30, 30, 30, 30
- b) Find 4 batting scores that have a mean score of 30 and a standard deviation of 10.
20, 20, 40, 40
- c) Find 4 batting scores that have a mean score of 30 and a standard deviation of 20.
10, 10, 50, 50
- d) Explain the method you used to arrive at your answers above.
The sum of the scores must be 120 (for same mean score). Also, the size of each deviation must be the same.
- e) Find 4 batting scores (between 0 and 100 inclusive) that have the largest possible standard deviation.
0, 0, 100, 100; SD = 50
- f) Find 4 batting scores (between 0 and 100 inclusive) that have the smallest possible standard deviation.
Any 4 identical scores (e.g. 23, 23, 23, 23) SD = 0
- g) Think about your answers to questions e and f. In your own words explain how the standard deviation is related to the spread of a data set.
The larger the standard deviation, the greater the variation in the data set; the smaller standard deviation, the smaller the variation in the data set.

Part 3: Changing the data

Consider the original 4 scores for Jamal. If each score had been 20 runs more, it is clear that his mean score (his 'batting average') would be higher, but by how much? Also, how would such an increase in each score affect the value of the standard deviation?

To do this with the aid of the spreadsheet template, we will add a formula in cell D5 that calculates the mean score.

- Move to cell D5.
- Type the formula **=approx(mean(score))**.

	A score	B dev	C sqdev	D
=		=score-m	=dev^2	
1	29	-1	1	meansq...
2	35	5	25	9.
3	29	-1	1	stdev
4	27	-3	9	3.
5				30.
D5	=approx(mean(score))			

Now use your spreadsheet template to help answer these questions.

Question 9.

What effect does adding 20 runs to each score have upon

- the mean score?
- the standard deviation of the score?
- Try repeating this by adding another 20 runs to each score.
- Why do you think this happens?

Finally, we investigate what happens to the mean and standard deviation if we multiply each score by a given number (e.g. $\times 2$ or doubling). What effect does this have on the mean and standard deviation?

Use your spreadsheet template to help answer the following questions.

Question 10.

What effect does doubling each of Jamal's original scores have upon the

- the mean score?
Mean score is doubled.
- the standard deviation of the score?
Standard deviation is doubled

Experiment with some other values (e.g. tripling each score), and then copy and complete the following table.

	Original scores	Original scores x 2	Original scores x 3	Original scores x 0.5
Score 1	29	58	87	14.5
Score 2	35	70	108	27.5
Score 3	29	58	87	14.5
Score 4	27	54	81	13.5
Mean	30	60	90	15
Standard deviation	3	6	9	1.5

Question 11.

In your own words, summarise what happens to the values of the mean and standard deviation when each score is multiplied by a constant factor. Suggest a reason why this might happen.

Both the mean and the standard deviation are also multiplied by that constant factor. The standard deviation increases by the factor, since the difference of each score from the mean increases by that factor.

Challenge

What would happen to the values of the mean and standard deviation if you were to double each score and then add 10? Try to do this by using the results from Part 3, and then check to see if you are correct. Generalise your result: if x represents the scores, then what will happen to the mean and standard deviation if each score is multiplied by a and then b is added (i.e. change x to $ax + b$)?

Teacher notes

- In this task, the construction of the standard deviation as a measure of spread is introduced. The focus is on its use as a descriptive measure of spread, and so the formula used is

$$\text{Standard deviation} = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

- In Part 1, the concept of deviation from the mean, and the problem of using signed deviations to calculate a measure of the spread are introduced. The development of the idea of an unsigned deviation (by squaring and then taking the positive square root of the deviations) is introduced as a more useful measure of spread.
- Students may not have been exposed to the formal notation of a statistical variable, and so this could be discussed: ‘We are interested in the central value of the variable, and how well spread out are the values of this variable.’
- Note the use of the “approx.” command. This is used to prevent the TI-Nspire CAS giving exact answers in fractional form for the mean squared variation and the standard deviation, which might obscure the idea for some students.
- Part 2 explores how the four values could be chosen to achieve a particular value of a standard deviation. This is free exploration—it would be a good outcome if the students can recognise how the standard deviation is affected by values that are very close or identical, as well as values that are well spread out.
- Part 3 highlights the invariance of the standard deviation when a constant amount is added to a statistical variable, but not when the variable is multiplied by a constant factor. This notion is often taken up more formally in senior mathematics courses.

$$- \text{SD}(X + b) = \text{SD}(X)$$

$$- \text{SD}(aX) = a\text{SD}(X)$$

$$- \text{SD}(aX + b) = a\text{SD}(X)$$

- In senior mathematics courses where the standard deviation of a sample is used as an estimate of the population standard deviation, it is usual to use $n - 1$ rather than n in the definition.