

1. Explanation:

Statements (1) and (2) are both insufficient by themselves because they tell us nothing about the data in the other months. We are left with either (C) or (E) as potential answers. Since we know the standard deviation for the two halves of the year, does that mean we can calculate the standard deviation for the whole year? Not necessarily. If the means of the two sets are different, then the mean of the combined set of earnings numbers will have to be calculated, and the standard deviation of the combined set depends on this new mean, which is impossible to determine.

Answer:

The answer, therefore, is (E).

2. possibility (a) - literal interpretation of what you wrote:

IF you're trying to say that the standard deviation is the *same number*, regardless of the number of consecutive integers in the set, then that's wrong.

you could use the formula you wrote to prove this, but here's a more conceptual way to think about it:

say we have a set of 7 consecutive integers.

then these integers are 3 less than, 2 less than, 1 less than, equal to, 1 more than, 2 more than, and 3 more than the mean.

now let's say we have a set of 9 consecutive integers.

then that's the same as the set of 7 consecutive integers - except we've added numbers that are 4 less than and 4 more than the mean. since these new numbers are farther from the mean than any of the pre-existing numbers, it follows that the standard deviation must be a *bigger number* once we've added those numbers.

possibility (b)

IF you're trying to say that it's good enough to be given the EXACT NUMBER of consecutive integers in the set, REGARDLESS of what that number actually is, then you're right.

in other words:

any set of, say, 7 consecutive integers must have the same standard deviation as any other set of 7 consecutive integers.

(the reason is because, as mentioned above, any such set consists of numbers that are 3 less than, 2 less than, 1 less than, equal to, 1 more than, 2 more than, and 3 more than the mean, regardless of the actual value of the mean.)

So the answer is A

3. the poster above did a good job of explaining the significance of the symmetry statement, so i don't need to rehash that. thanks, poster above.

what the poster above didn't write is that *it actually makes no difference at all that 'd' is the standard deviation* in this problem. 'd' could be a completely random number and the problem would still work out the same way.

they're writing the problem with 'd' as the standard deviation for at least one, and probably both, of the following reasons: (a) to mess with your head, and (b) because IF 'd' IS the standard deviation, then the 68% and 16% take on a special significance (which is irrelevant here, but of tremendous importance in statistics).

but:

just think about SYMMETRY.

think of 'm' in the middle of a number line. then 'm + d' is just as far to the right of it as 'm - d' is to the left of it.

so:

(1) 68% of the stuff lies between m - d and m + d.

this means the other 32% (= 100% - 68%) of the stuff lies outside those boundaries.

because of the symmetry in the problem, this means that 16% of the stuff is to the left of m - d, and 16% of the stuff is to the right of m + d.

answer = 16%

sufficient

(2) by symmetry, the amount of stuff to the left of m - d must be the same as the amount of stuff to the right of m + d (because those two regions are mirror images of each other under the symmetry in the problem).

thus, answer = 16%

sufficient

answer = d

5. You never need to calculate SD on the GMAT. In DS, we're just seeing if we have enough info to calculate.

In this question, if we know what the sets are, we can certainly calculate SD.

We already know 3/5 numbers on each list, so we're not far off. We also know the average of each list, and therefore the sum (= avg * # of terms).

(1) S contains 25. This allows us to determine the 5th member of S, but we don't know enough about T.

(2) T contains 45. This allows us to determine the 5th member of T, but we don't know enough about S.

Together: We know the full sets for S and T. If we know all the terms in a set, we can calculate ANYTHING about that set, including SD. If we can calculate both SDs, we can certainly answer the question: choose (c).

6. Statement I

all integers are positive

T = [15,15,15]
SD = 0

T = [5,15,25]
SD = positive

Insufficient.

Statement II

T = [5]
SD = 0
T = [25]
SD = 0

Sufficient.

Hence B

7. Book Answer: (C)

"Statement (1) is not sufficient because it tells you the mean, but nothing about the other numbers in the set. You could Plug In two sets of numbers for the set and get two different standard deviation. Eliminate choices (A) and (D). Statement (2) is also insufficient. Again, you could Plug In two sets of numbers to get two different standard deviations. Eliminate choice (B). Combining the two statements tells you that the numbers in set S are consecutive, because in a consecutive set of numbers, the mean equals the median. If the numbers are consecutive, even integers and the mean is 4, you know that the standard deviation is 2."

Standard deviation :

The square root of the sample variance of a set of N values is the sample standard deviation

$$s_N = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} .$$

where x_i is ith member of set

\bar{x} with bar over head is mean of the set.

the expression above is taken off from <http://mathworld.wolfram.com/StandardDeviation.html>

since the difference between the mean , 4 in this case, and other consecutive numbers, like for example the set be [-4,-

2,0,2,4,6,8,10,12] will be a multiple of 2, not exactly 2. when we square them , sum them and divide the sum by the number of members, which is 9

we get Std Dev = $\sqrt{240/9}$ = 5.16

which is not 2 definitely.

8. IMO it is D

statement1)

The range of this data set is 0.

Range = highest number in the set - lowest item. The range will only be 0 when all the items are same. Hence third item = first item = 3....sufficient

statement2)

The standard deviation of this data set is 0.

standard deviation = $\sqrt{\text{variance}}$

variance = summation $(X-M)^2/N$...N - number of items. This will only be 0 when all the items are equal to the mean. This is only possible if the mean is same as all the items. Hence all items will have to be 3. Sufficient.

IMO D

10. That's easy. Knowing the average value doesn't tell us deviation from this mean which is reqd for SD. 1) insufficient
2) directly tells us that SD is 0 as all the values are same. hence B is sufficient!

11.

Std. Dev. = $\sqrt{\text{sum for all terms of } (X - \text{mean})^2/n}$, where X is each term and n is the number of terms.

This is probably not the best way to represent it (the limitations of posting on a forum, alas!) but for our purposes it is OK. On the GMAT it is more important to understand what std. dev. represents than to know how to calculate it.

Qualitatively, std. dev. is a measure of the spread of the data.

Harish, the source of your confusion is your statement "I know that the standard deviation of the sample **doesn't change** if we add or subtract the same **constant value** to the sample values." That is only true if all of the samples have the same quantity to begin with (std. dev. = 0)!

The more accurate statement would have been "The standard deviation of the sample **changes by a known factor** if we add or subtract the same **percentage** to each of the sample values." If the samples each decrease by 30%, the mean decreases by 30%, and the $(X - \text{mean})$ decreases by 30% for each term. You don't really have to complete the calculation to see that the resulting std. dev. will be smaller than the original 10 by some factor (I believe the result would be 7, but you can check my math).

you can PICK ONE change to make:

EITHER

* change the numbers by the same *percentage*,

OR

* change the numbers by the same *numerical amount*.

you *cannot* do both of these at once, unless the two numbers are the same to start with, because the same *absolute* change will constitute a different *percentage* of each of the two starting numbers.

in the example you've quoted, 2 is 20% of 10, so you're decreasing the original value of 10 by 20%. by contrast, 2 is 40% of 5, so you're decreasing the original value of 5 by 40%.

if you were to decrease by identical percentages, then you'd either do 10 - 4 and 5 - 2 (40% decrease each), or 10 - 2 and 5 - 1 (20% decrease each).