First, here are some challenging practice questions:

1) Set S has a mean of 10 and a standard deviation of 1.5.  We are going to add two additional numbers to Set S.  Which pair of numbers would decrease the standard deviation the most?

(A) {2, 10}  
(B) {10, 18}  
(C) {7, 13}  
(D) {9, 11}  
(E) {16, 16}

2) Set Q consists of the following five numbers: Q = {5, 8, 13, 21, 34}.  Which of the following sets has the same standard deviation as Set Q?

I. {35, 38, 43, 51, 64}  
II. {10, 16, 26, 42, 68}  
III. {46, 59, 67, 72, 75}

(A) I only  
(B) I & II  
(C) I & III  
(D) II & III  
(E) I, II, & III

3) Consider the following sets:  
L = {3, 4, 5, 5, 6, 7}  
M = {2, 2, 2, 8, 8, 8}  
N = {15, 15, 15, 15, 15, 15}

Rank those three sets from least standard deviation to greatest standard deviation.

(A) L, M, N  
(B) M, L, N  
(C) M, N, L  
(D) N, L, M  
(E) N, M, L

Do these three questions make your head spin?  You have found a good blog article to help you!  Explanations to these will appear at the end.

Spread

When we are summarizing a list of numbers, typically we want to know the **center** and the **spread**.  (If we are doing an advanced analysis, we would also want to know the shape of the distribution: that can come into play in [IR questions](http://magoosh.com/gmat/2012/gmat-integrated-reasoning-ebook/).)

The two most typical measures of center are [mean and median](http://magoosh.com/gmat/2012/common-gmat-topic-descriptive-statistics/).  Center gives us an idea of where the middle of the distribution of numbers falls.

Measures of spread give us an idea of the spacing of the numbers, how much they are “spread” out from each other.  A relatively crude measure of spread is the [range](http://magoosh.com/gmat/2012/common-gmat-topic-descriptive-statistics/), which really only tells us about the extreme high and the extreme low, not all the data points in the middle.  A more sophisticated measure of spread is the standard deviation.

Standard Deviation

Every list of numbers has a mean.  Therefore, every number on the list has a **deviation from the mean**: that is how far that number is from the mean.

deviation from the mean = (value) – (mean)

Technically, numbers below the mean have a negative deviation from the mean, and numbers above the mean have a positive deviation from the mean.   In the list {2, 4, 6, 8, 10}, the mean = 6, so 8 has a deviation from the mean of +2, and 2 has a deviation from the mean of -4.  So, parallel to this first list is a second list, the list of deviations from the mean.  (It’s a good exercise to convince yourself why this second list always has a mean of zero.)

*Here is the technical procedure for calculating the standard deviation*.  We already have List #1, original data set, and List #2, deviations from the mean for each value in List #1.  Now, List #3 will be the List #2 squared — the squared deviations from the mean.  This is the list we average: that average is something called the “variance.”  Then, to undo the effects of squaring, we take a square root, and that final answer is the standard deviation.  The OG explains this procedure in the Math Review.  If you understand and remember this, great, but chances are good that you don’t need to know it in all its gory detail if you know the rough and ready facts below.

Rough and ready facts about standard deviation

1) The standard deviation gives us an estimate of the size of a typical deviation from the mean.  It’s a way of “averaging” the deviations from the mean, though it is not strictly the mean of that list.

2) If every element in the data set is equal, they all equal the mean, each deviation from the mean is zero, and the standard deviation is**zero**.  This is the **lowest possible standard deviation** for any set to have. (That’s an excellent GMAT shortcut to know!)

3) If you add the same number to every number on the list, or if you subtract the same number from every number on a list, or if you subtract each number on the list from the same number, all of the new lists produced would have exactly the same standard deviation as the original.  Addition and subtraction slides values up and down the number line, but does not change any of the spacing between the numbers.

4) If you multiply the numbers on a list by any values (other than ±1), or if you raise the numbers on a list to a power, that *always* changes the standard deviation.  Multiplying changes the spacing on the list.  In particular, if you multiply each number by k, then you multiply the standard deviation by |k|.

5) If all the numbers on the list are *the same distance from the mean*, that distance *is*the standard deviation.  For example, in the set {17, 17, 17, 23, 23, 23}, the mean = 20, and each number is exactly 3 units from the mean, so the standard deviation is 3.

6) If you do anything that “bunches the numbers together”, that decrease the standard deviation.  If you do anything that “pulls the numbers further apart”, that increase the standard deviation.

7) If you include new numbers in the set — that is tricky, because adding in most numbers will *change the mean of the entire set*, which will change the deviation from the mean for each number on the list, which changes the standard deviation.  If you include an additional number or a few additional numbers that are far away from the other numbers, this inclusion will wildly increase the standard deviation.

8) If you include two new numbers that are symmetrical around the mean, then that will not change the mean.  If the distance of these two numbers from the mean is greater than the standard deviation, adding them will increase the standard deviation (there’s a larger “average” distance from the mean).  If the distance of these two numbers from the mean is less than the standard deviation, adding them will decrease the standard deviation (there’s a smaller “average” distance from the mean).

9) This is an extreme instance of the last case discussed in the previous point.  If you include two new numbers equal to the mean (and therefore, with a deviation from the mean of zero), of course that decreases the standard deviation, but we can say more than that.  Of all possible new numbers you could include in a set, the new numbers that will ***most decrease*** the overall standard deviation of the set are new entries equal to the mean.  That is the single most efficient way to decrease the standard deviation of a set by including new entries to the list.

I realize that’s a great deal of information.  The more you understand how standard deviation works, the more you will understand the interconnection of these “rough and ready” facts, which will make the entire list easier to remember.

At this point, you may want to go back to the three practice questions at the beginning of this post, and see if you have any insights.

Practice problem solutions

1) This is a very tricky problem.  Starting list has mean = 10 and standard deviation of 1.5.

(A) {2, 10} — these two don’t have a mean of 10, so adding them will change the mean; further, one number is “far away”, which will wildly decrease the mean, increasing the deviations from the mean of almost every number on the list, and therefore increasing the standard deviation.  **WRONG**

B. {10, 18} — these two don’t have a mean of 10, so adding them will change the mean; further, one number is “far away”, which will wildly increase the mean, increasing the deviations from the mean of almost every number on the list, and therefore increasing the standard deviation.  BTW, (A) & (B) are essentially the same change — add the mean and add one number eight units from the mean.  **WRONG**

C. {7, 13} — centered on 10, so this will not change the mean.  Both of these are a distance of 3 units from the mean, and this is larger than the standard deviation, so it increases the size of the typical deviation from the mean.  **WRONG**

D. {9, 11} — centered on 10, so this will not change the mean.  Both of these are a distance of 1 units from the mean, and this is less than the standard deviation, so it decreases the size of the typical deviation from the mean.  **RIGHT**

E. {16, 16} — these are two values far away from everything else, so this will wildly increase the standard deviation.  **WRONG**

Answer = **D**

2) Original set: Q = {5, 8, 13, 21, 34}.

Notice that Set I is just every number in Q plus 30.  When you add the same number to every number in a set, you simply shift it up without changing the spacing, so this doesn’t change the standard deviation at all.  Set I has the same standard deviation as Q.

Notice that Set II is just every number in Q multiplied by 2.  Multiplying by a number *does* change the spacing, so this *does* change the standard deviation.  Set II does **not**have the same standard deviation as Q.

This one is very tricky, and probably is at the outer limit of what the GMAT could ever expect you to see.  The spacing between the numbers in Set III, from right to left, is the same as the spacing between the numbers in Q from left to right.  Another way to say that is: every number in Set III is a number in Q subtracted from 80.  Again, would be very hard to “notice”, but once you see that, of course adding and subtraction the same number doesn’t change the standard deviation.  Set III has the same standard deviation.

The correct combination is I and III, so the answer is **C**.

3) OK, well first of all, set N has six numbers that are all the same.  When all the members of a set are identical, the standard deviation is zero, which is the smallest possible standard deviation.  So, automatically, N, must have the lowest.  Right away, we can eliminate (A) & (B) & (C).  In fact, even if we could do nothing else in this problem, we could guess randomly from the remaining two answers, and the odds would be in our favor.  See [this post](http://magoosh.com/gmat/2012/guessing-and-skipping-strategies-for-the-gmat/) for more on that strategy.

Now we have to compare the standard deviations of Set L and Set M.  In Set L, the mean is clearly 5: two of the entries equal 5, so they have a deviation from the mean of zero, and no entry is more than two units from the mean.  By contrast, in Set M, the mean is also 5, and here, every number is 3 units away from the mean, so the standard deviation of M is 3.  No number in Set L is as much as 3 units away from the mean, so whatever the standard deviation of L is, it absolutely must be less than 3.  That means, Set L has the second largest standard deviation, and Set M has the largest of the three.  N, L, M in increasing order.  Answer = **D**.