1. D

From the given #s it is clear 4 is mean(average) of both sets.
If the **variations** of the sets are equal or close then their Stand Deviations are equal/close to each other too.

Variation of 1-set (5 #s) is 40/5=8
so Varition of set2 (7 #s) must be around 8 too, thus (40+ x)/7= around 8

x must be close to 16 so only D satisfies this value: (4-2)^2 +(4-6)^2=8
U can check others:
A) (4+1)^2+(4-9)^2=50
B) (4-4)^2+(4-4)^2=0
c) (4-3)^2+(4-5)^2=2
D) (4-2)^2 +(4-6)^2=8
E) (4-0)^2+(4-8)^2=32

This is a good explanation. You can also do it by looking at the numbers and using your common sense about standard deviation.

The average of the first list is 4 and the number deviate from 4 evenly. what i mean by that is, the next set of numbers (2,6) are both 2 away from 4, and the next set after that (0,8) are both 4 away from 4.

To have a similar deviation, we want the next set to look similar.

-1 and 9 will stretch the outter limits of the list, so that will increase the standard deviation significantly.
4 and 4 will add too much weight to the center of the list. It'll decrease the standard deviation.
3 and 5 may be right
2 and 6 may be right
0 and 8 will add weight to the outter limits again, and stretch the deviation.

So it's a toss up between (3 and 5) and (2 and 6). And my educated guess is on 2 and 6, since basically comes in right in the center of the previous standard deviation, it should change it the least.

For the record, I have never taught the standard deviation formula to any student of mine. I find it to be unnecessary for the GMAT when good, conceptual thinking can get you through.

2. ok, so "standard deviation" measures the SPREAD of the data around the average. in other words, **the farther the data are from the average, the greater the s.d.**

in other words:

* if you **ADD data that are FARTHER FROM THE AVERAGE** than are the existing data, then the **s.d. will INCREASE**.

* if you **ADD data that are CLOSER TO THE AVERAGE** than are the existing data, then the **s.d. will DECREASE**.

this means that you have to add data that are CLOSER to the average (i.e., 6).

the only way to *ensure* this is to add actual 6's, because you have no idea how far the existing points are from 6; for all you know, the entire set could be a bunch of 5.9999's and 6.0001's. (all you know is that at least some of the data points are non-6's, because otherwise the standard deviation would be 0 and couldn't be decreased.)

therefore, **(e)**.

3. Let d stand for Standard Deviation, M - for mean. We can construct a system of equations:

$$\begin{cases} M-2d=58 \\ M+3d=98 \end{cases}$$

After you subtract the first equation from the second, you get this equation:

$$5d=40$$
$$d=8$$

After plugging the value of d in the first equation, you get the value of M:

$$M=58+2*8=74$$

I hope it helps you.

4. Look for range and # of elements in the set.

A set with higher the range and fewer the number of element has the higher SD. i.e. A.

5. (D) I and II only

Anything added/deducted to the set elements or the set elements deducted from anything results in no change in SD.

I. Deduct 2 from each of the elements in set result I. i.e. {r-2, s-2, t-2}
II. Deduct each set elements from s. The new set elements in II i.e. {0, s-t, s-r} result.
III. Taking the absolute value of the set elements is not the same as deducuting or adding the same. This act would not change the SD if all set elements have the same sign (+ve or -ve).

Suppose s = 5 and r = 6 and t = 7, {|r|, |s|, |t|} and {s, r, t} have same SD.
If s = -5 and r = -6 and t = -7, {|r|, |s|, |t|} and {s, r, t} have same SD.
If s = -5 and r = 6 and t = 7, {|r|, |s|, |t|} and {s, r, t} have different SD.

III is not always a true case.

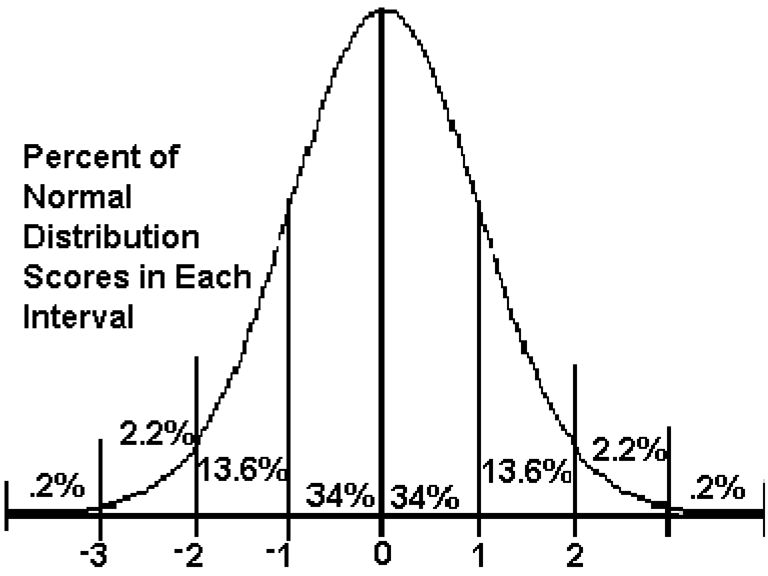
6. D

The prompt says that 68% of the population lies between m-d and m+d.

Thus, 32% of the population is less than m-d or greater than m+d.

Since the population is symmetric, half of this 32% is less than m-d and half is greater than m+d.

Thus, (68+16)% or (100-16)% of the population is less than m+d.



7. The standard deviation hinges on the sum of the squared differences between the values and the mean. These calculations therefore boil down to:

Set A: (1 – 3)^2 + (3 – 3)^2 = 4
Set C: (2 – 3)^2 + (2 – 3)^2 = 2

Since the sum of the squared differences is larger in A, A has the larger standard deviation and is therefore the third largest overall.

The correct answer is **A.**

8. In general, you should definitely know how to handle both average (mean) and median. Standard deviation is necessary if you want a high score.

Some of the discussion in the explanation for this problem is basic to understanding what median and mean represent. For example, if the median of a set of numbers is greater than the mean of that same set - this just means that the numbers below the median must be farther from the median than the numbers above - eg 1, 3, 4, 16, 17, 18, 19, the median is 16 and the mean is 11.1. If the median is smaller than the mean, then the numbers above the median must be farther from the numbers below - just the opposite of the previous example.

Also know that there are some relationships between mean and median when you have a set of consecutive integers (must be consecutive). If you have a set of consecutive integers, then you can find the average by just averaging the first and last terms. For example, 1, 2, 3, 4, 5. For the average, I can just take (1+5)/2 = 3. If the set has an odd number of consecutive terms, then the median will also equal the mean. In that last example, 3 is both the median and the mean.

The rule given for a "composite set" in this explanation is a more obscure rule, however. Only worry about remembering that if you want a 700+ score. (Notice that this problem is labeled 700-800; it's a very hard problem.)

Given x>y and L = M

this means that while set A is not equally spaced, set B is equally spaced..

getting to the statements

1) Z > N ?

This definitely does not have to be true...

when we say that set B is equally spaced... we dont mean that the set will be consecutive integers e.g 1 ,2 ,3 ...

it could equally spaced multiples of 5 or 7 ....

and as for set A, it could have just 3 set of numbers e.g 1 3 4 where the standard dev would be less...

Hence we can have a case where SD of set B > SD of set A or N > Z

take for example set B as 100 200 300 500 600 and set A as 1 3 4

Statement 2) R > M

R > M can only be decided if we have a relation between Y & M

as adding a number greater than the mean will increase the overall mean

and subtracting a number lesser than the mean, will reduce the overall mean...

since we dont know what Y and M are... or the relation between them, we cant deduce anything...

Statement 3) Q > R

Again, this depends on the actual elements of the set

By the similar logic, I have given for statement 2, u can have different possibilities for Q and R

none can be said for sure...

i would choose **E:none**

9. Since E is the collection of four numbers and the greatest difference between any two integers in E is 4 so the numbers can be 1,3,3,5 or 1,1,3,5 or 1,3,5,5 or 3,5,5,7 or 3,3,5,7 or 3,5,7,7

10. If the SD is 3.0 and we're looking for the answer choice that is 2.5 stdevs from the mean, then we're looking for an answer choice that is less than 20 - 3*2.5 = 12.5 or greater than 20 + 3*2.5 = 27.5.

Only **[A]** fits the bill here.

11. Mean is 13.5

SD = 1.5

2sd = 2* 1.5 = 3

We need to find the value exactly 2 sd less than the mean so we need to find 13.5-3

If we had to find the value exactly 2 sd greater than the mean we would do 13.5+3

In general :

Points k Sd above the mean : x+k*SD

points k sd below the mean : x-k*SD

10.5, one SD away means +/- SD so 2 less means: 13.5-3 **A**