We will work our way through the concepts of Standard Deviation (SD) today. Let's take a look at how you calculate standard deviation first:

$$SD = \sqrt{\frac{\sum(Ai - Aavg)^2}{n}}$$

Ai – The numbers in the list

Aavg – Arithmetic mean of the list

n – Number of numbers in the list

Say you have 3 numbers : 11, 13 and 15. Their standard deviation is the "square root of the average of their squared deviations from the arithmetic mean." Let's see what we mean by this.

Mean of 11, 13 and 15 is 13.

$$\sum(Ai - Aavg)^2 = (11 - 13)^2 + (13 - 13)^2 + (15 - 13)^2 = 8$$

$$SD = \sqrt{\frac{8}{3}}$$

Focus on these words: "deviations from mean"

The important point to note is that SD is a measure of dispersion or deviation from the mean (the mean is approximately the middle of the list if there are no outliers). In other words, SD is a measure of whether the numbers are very far away from the mean or close together. Since GMAT isn't calculation intensive, you probably won't need to calculate the actual SD in the test. The calculations are shown here only to illustrate the concept. But you must have a feel for how the numbers are distributed around the mean and what that implies for the SD.

Your statistics book explains how to visualize SD using the number line in detail, therefore, I am not going to delve deep into it but will quickly recap so that we can move ahead. Recall that if you plot the numbers on the number line, it gives you a sense of how far the numbers are from the mean. The farther the numbers, higher is the SD.

Let's check out a few different cases to internalize the SD concept. Do not calculate anything in these questions. Just look at the number line for each case and figure out whether it makes sense to you.

Question: Which set, S or T, has higher SD?

Case 1: S = {3, 3, 3} or T = {0, 10, 20}

Case 2: S = {3, 4, 5} or T = {5, 6, 7}

Case 3: S = {3, 4, 5, 6} or T = {2, 3, 4, 5, 6, 7}

Case 4: S = {1, 3, 5} or T = {1, 1, 3, 5, 5}

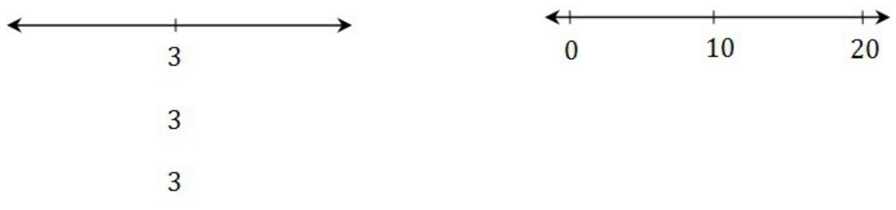
Case 5: S = {1, 3, 5} or T = {1, 3, 3, 5}

Case 6: S = {6, 8, 10} or T = {12, 16, 20}

Case 7: S = {6, 8, 10} or T = {3, 4, 5}

Let me represent the first four cases on the number line. Check them out and then think which set should have the higher SD.

Case 1: S = {3, 3, 3} or T = {0, 10, 20}



Case 2: S = {3, 4, 5} or T = {5, 6, 7}

Case 3: S = {3, 4, 5, 6} or T = {2, 3, 4, 5, 6, 7}

Case 4: S = {1, 3, 5} or T = {1, 1, 3, 5, 5}

Let’s discuss each of these four cases now.

Case 1: S = {3, 3, 3} or T = {0, 10, 20}

T has higher SD. We will obtain the SD of T by calculating as shown in the example above. But we don’t really need to calculate it because we see that for set S, SD = 0. Each number is at the mean and hence has 0 deviation from the mean. Since SD cannot be negative, whatever the SD of T, it will be higher than the SD of S which is 0.

Case 2: S = {3, 4, 5} or T = {5, 6, 7}

Both sets have the same SD. We can see from the number line that they are equally dispersed around their respective means.

Case 3: S = {3, 4, 5, 6} or T = {2, 3, 4, 5, 6, 7}

Set T has higher SD. T has two extra numbers which are farther from the mean. Hence these 2 numbers will add to the total deviation. (There is a caveat here which we will discuss next week.)

Case 4: S = {1, 3, 5} or T = {1, 1, 3, 5, 5}

T has higher SD. It has two extra numbers far from the mean. (There is a caveat here too!)

What do you think about cases 5, 6, and 7? I will give you the answers to these three cases next week!