LRNA/PSME 95 Final Project
Yi Tao Zhao, Yumeng Yang, Zechang Li

HOW TO INTRODUCE STANDARD DEVIATION

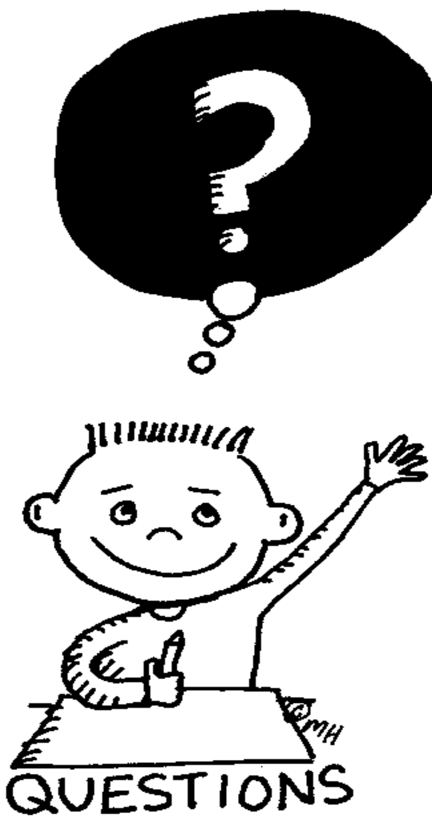
Standard deviation is one of the most essential fundamental concepts in statistics class. And it is also very common for tutees to have trouble with, which causes them to feel confused about the materials that they learn later on, like the analyses of distribution and z-score.

Description of tutee's difficulty

The difficulty that tutees complain the most is that they feel uncomfortable to purely memorize the formula looks so complicated. In another world, they are scared by that complex-looking formula which involves summation notation, square and square root at the same time. They often feel unconfident with themselves at the very first place and expect their magic calculator could do find result for them.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}.$$

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$

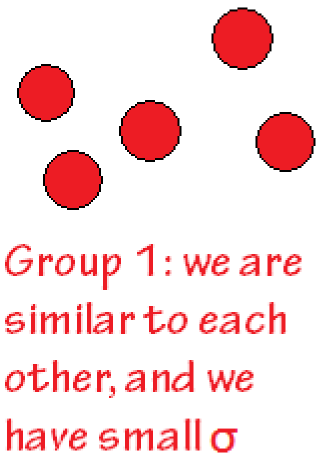


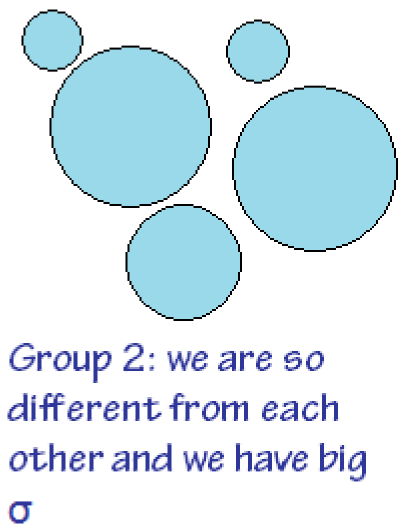
Analysis of Concept

-What is a Standard Deviation?

Standard deviation is a widely used measurement of variability or diversity used in statistics and probability theory. It shows how much variation or 'dispersion' there is from the 'average' (mean, or expected/budgeted value). A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data is spread out over a large range of values.

A simple example is helpful here:





For instance, we have two groups of data such as

GroupA = {1,1,1,1,1}   GroupB = {0,100,0,100,50}

Find average for both of them

$\text{Xbar}_{\text{gropuA}}=1$ $\text{Xbar}_{\text{groupB}}=50$

Then we could let tutees try to plug the numbers into the formula at first. Although our purpose is to let them understand the concept, especially for statistics, plug and chart method sometimes is helpful for them to get a preview about the knowledge.

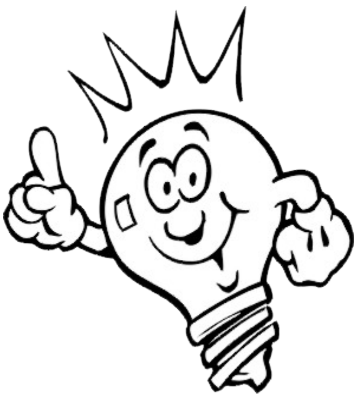
$\sigma_{\text{groupA}}= 0$ $\sigma_{\text{gropuB}} = 100$

Accordingly we should easily see that when data are distinct from each other we will get a relative large standard deviation, whereas, if the data members are similar the standard deviation will turn out to be small.

Making connection:

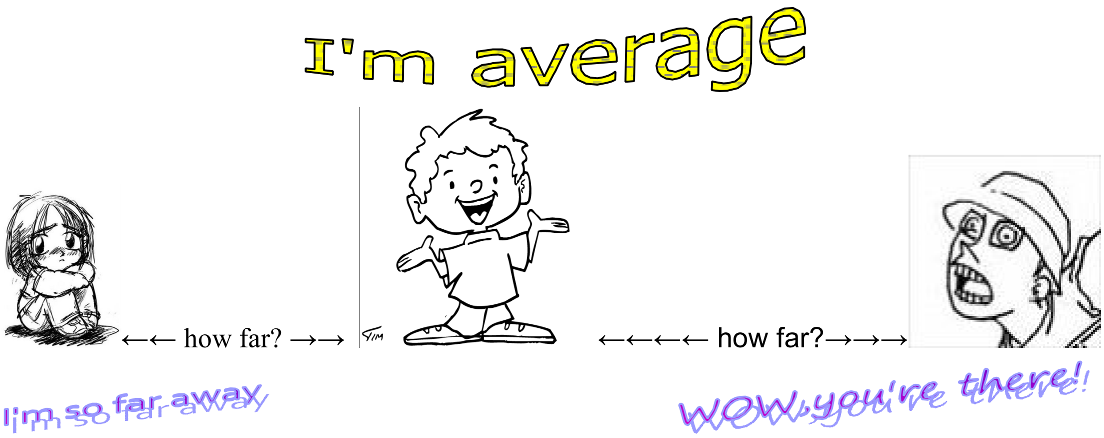
As a brand new mathematical measurement for most students, standard deviation does not appear as so familiar as average to them. But actually they are closely associated with each other, so that our tutors should emphasize the connection in between to help them understand it.

Instead of the complicated mathematical expression, we could describe the Standard Deviation as the average distance from each data to the mean value.



Tips

It's helpful to have some little activities here, to place some object, like an eraser, on the table as the mean value. And put some little objects, like pen tops surround. It's better to set all the object in a line. These are representing the data, which have different distances away from the average.



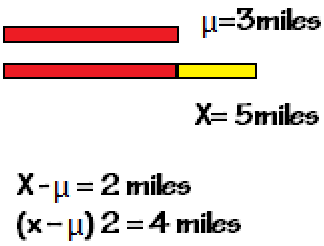
Therefore to find out the average distance we need to add all these little pieces of distances together.

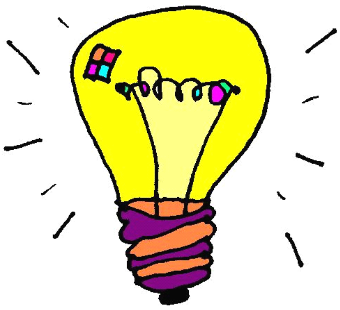
To find the distance we use each data to subtract the average

$$x-\mu$$

Meanwhile, in calculation some data is less than the average and some other data is larger than the average. If we just want to know the length of the distance, we don't need to worry about if there is a negative sign; therefore take the square of the subtraction will make all the distance value positive.

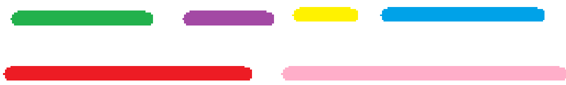
Thus we do $(x-\mu)^2$ and its must be positive.





Then it's the time to find the average distance, we sum all the little parts up, and divide by the amount of data.

$$\frac{(x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2+(x_4-\mu)^2+\ldots..}{N}$$

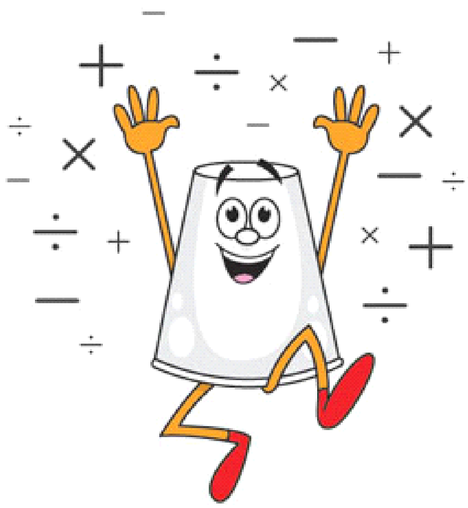


We obtained these little pieces of (x-u)^2 and sum them up.

Then devide the sum by the amount of the pieces, which is the amount of the number n.

Finally don't forget we have squared all the pieces of distances, so it's time for us to take the square root of that; then we shall have the standard deviation.

$$\sigma = \sqrt{\frac{1}{N}\left[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2\right]},$$



Note: There is one thing need to be noted. There is some little differences between **sample** and **population**, of which students should be aware. Population is the entire data for our target event, sample is only one portion of the population.

For example, if I do a survey on De Anza students' height, all De Anza students' height will be my population, while, if I just select 100 students from De Anza College, then that 100 people's height will be my sample.

And in terms of Standard Deviation, when looking for the standard deviation for sample, we have to divide the sum by (n-1) rather than n. Most time calculator will take care that. However it is essential to be aware of the differences between sample and population. Therefore as the concept in the beginning chapter, our tutor should mention this to tutees to give them a sense.