

S.4 Measures of Variation: Variance and Standard Deviation

S.4 Measures of Variation: Variance and Standard Deviation

We have looked at many statistical measures in this unit.

Some measures tell us how data points are similar to each other:

mean median mode

Some measures tell us how data points are spread out:

IQR range quartiles

An alternative measure of spread might take into account how far each data point is from the mean. Consider the following data:

$$49 \ 56 \ 55 \ 68 \ 61 \ 57 \ 61 \ 52 \ 63 \quad \overset{\text{mean}}{\bar{x} = 58}$$

We could calculate the "spread" by subtracting each data point from the mean, and then adding up all of those values ($x_i - \bar{x}$):

$$-9 + -2 + -3 + 10 + 3 + -1 + 3 + -6 + 5$$

When we add them all up, we get 0. In fact, we will always get this result. That's not very helpful! So we have to do something else to the values to remove all the negative values.

What we do is square those differences, add them up, and take the average.

$$81 + 4 + 9 + 100 + 9 + 1 + 9 + 36 + 25 = \frac{274}{9} = 30.4$$

This quantity is called the variance: $\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$

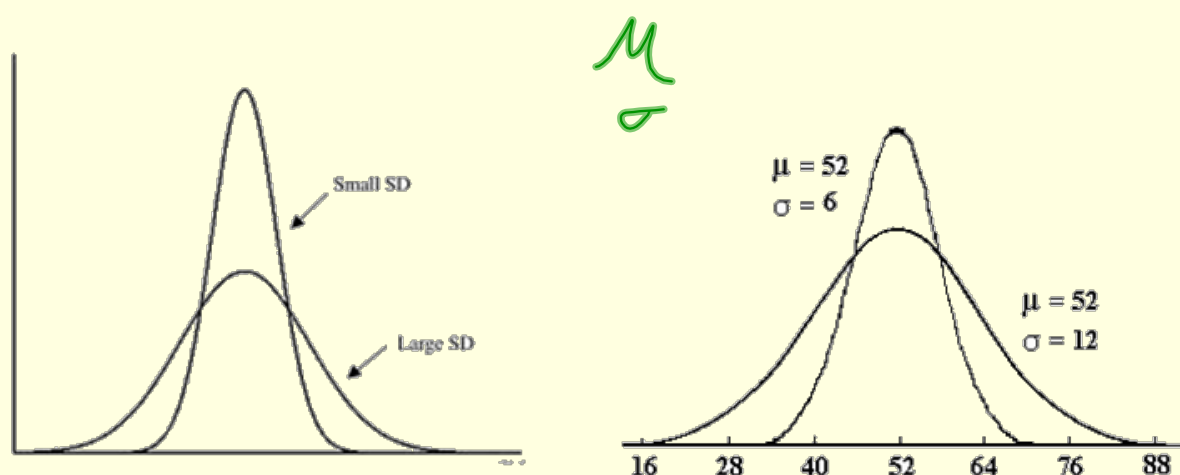
Since all the differences were squared, we often get a number much larger than the elements in the data themselves. Thus, it is hard to compare the variance to the original data. However, the square root of the variance brings the scale back down, and is much more helpful in relation to the original data. The square root of the variance is called the standard deviation.

$$\sqrt{30.4} \approx 5.51$$

$$\text{standard deviation} = \sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

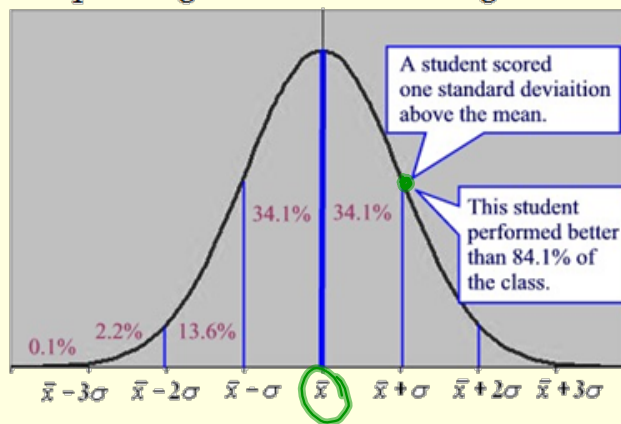
Small standard deviations indicate that the data in general is close to the mean. This might indicate that the data is stable or reliable. There is little variation.

Large standard deviations tell us the data is far from the mean. This might indicate the data is volatile or unreliable. There is a lot of variation.



**IB specifically states that you will not be expected to find the variance or standard deviation other than by a graphing calculator. HOWEVER, you do need to know the meanings of these measures, how the formulas are put together, AND that the standard deviation is the square root of the variance.

The standard deviation helps us figure out the meaning of data in relation to the mean.



The average on your Probability Test was 89.

The standard deviation was 3.5.

The scores on this test were close to to the mean.

Graph this information below.

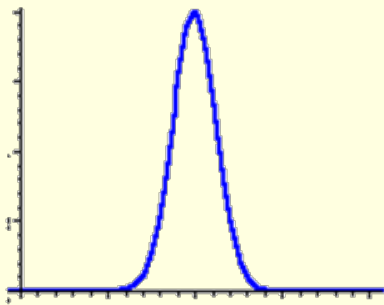
The average on your Functions Test was 79.

The standard deviation was 10.

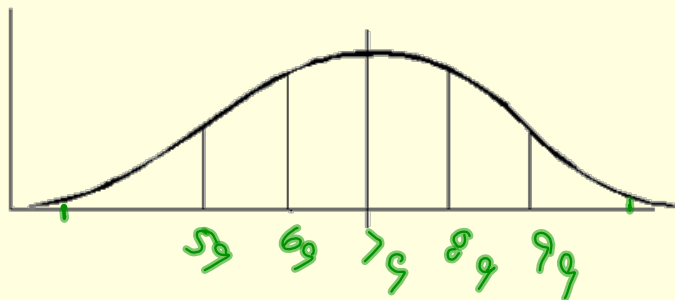
The scores on this test were far from the mean.

Graph this information below.

Probability Test



Functions Test



A student scored an 85 on both tests.

On which one did she score higher in relation to the mean?

```

1-Var Stats
 $\bar{x}$ =31
 $\Sigma x$ =372
 $\Sigma x^2$ =16910
 $Sx$ =22.11128877
 $\sigma x$ =21.6994725
 $n$ =12
    
```

Month	Selecta Disc	Discount Discs
January	100	80
February	30	1
March	30	70
April	30	2
May	21	70
June	23	1
July	21	1
August	21	70
September	24	3
October	21	70
November	30	2
December	21	2

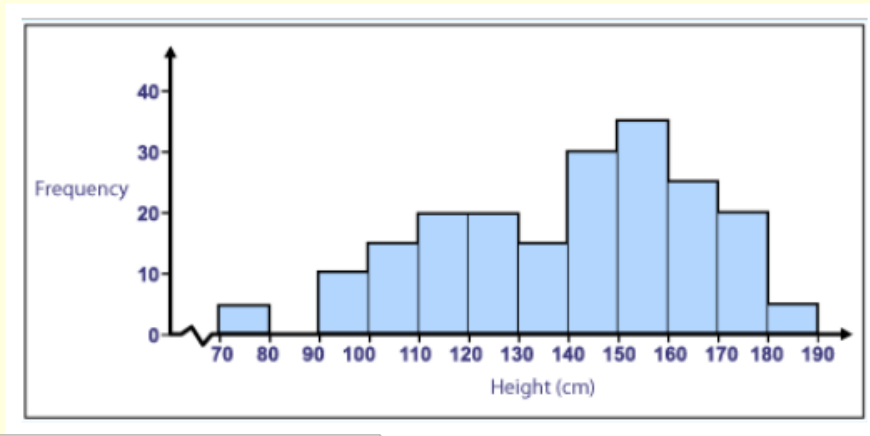
Selecta Disc's mean is 31. Discount Discs' mean is 31.

Selecta Disc's variance is . Discount Discs' variance is .

Selecta Disc's standard deviation is 21.2. Discount Discs' standard deviation is 34.8.

\bar{x} , μ

σ



L1 Height	L2 Freq.
75	5
95	10
105	14
115	20
125	20
135	14
145	30
155	35
165	25
175	20
185	5

The mean of the data is:

$$140.5$$

The median of the data is:

$$145$$

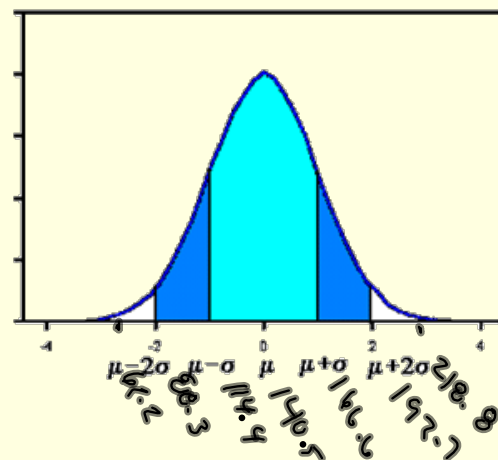
The mode of the data is:

$$155$$

The standard deviation of the data is:

$$\sigma = 26.1$$

Label the normal curve below:



Homework: WS S.4