- The **standard deviation** is a way of measuring **variation** in a population.

- **Standard deviation** is a means of determining the **reliability** within a sample population. This means how **close** each individual measurement is to the **average** of all the measurement.

  Consider the following results of two different tests

  72, 73, 74, 75, 76, 77, 78 Average = 75
  50, 60, 70, 80, 90, 100 Average = 75

  Each data set has the same average but clearly there is tremendous variation within each data set.

- **The formulae**

  - **Deviation**: $\delta_{x_n} = |x_n - \bar{x}|$

    where δ_{x_n} is the deviation for any given measurement, x_n any given measurement, and $\bar{x}$ is the mean of the sample.

	Data	$\delta_{x_n} = \lvert x_n - \bar{x}\rvert$
	72	$\delta_{x_1} = \lvert x_1 - \bar{x}\rvert$ $= \lvert 72 - 75\rvert$ $= 3$
	73	2
	74	1
	75	0
	76	1
	77	2
	78	3
$\bar{x}$	**75**	

- **Standard Deviation:** $S.D. = \sqrt{\frac{\sum_n^k (\delta_{x_n})^2}{k-1}}$

  where $\sum_n^k (\delta_{x_n})^2$ is the sum of all the squares of the deviations

	Data	$\delta_{x_n} = \lvert x_n - \bar{x} \rvert$	$(\delta_{x_n})^2$
	72	$\delta_{x_1} = \lvert x_1 - \bar{x} \rvert = \lvert 72 - 75 \rvert = 3$	9
	73	2	4
	74	1	1
	75	0	0
	76	1	1
	77	2	4
	78	3	9
$\bar{x}$	**75**		

$$S.D. = \sqrt{\frac{\sum_n^k (\delta_{x_n})^2}{k-1}}$$
$$= \sqrt{\frac{9+4+1+0+1+4+9}{7-1}}$$
$$= \sqrt{\frac{28}{6}}$$
$$= \sqrt{4.67}$$
$$= 2.16$$

- **The meaning of Standard Deviation**
  - **1 S.D.** means that about **2/3 (68%)** of the data falls within $\bar{x} \pm 1 S.D.$

    EX: 75 ± 2.16; this means that 2/3 of the test scores fall between 72.84 and 77.16

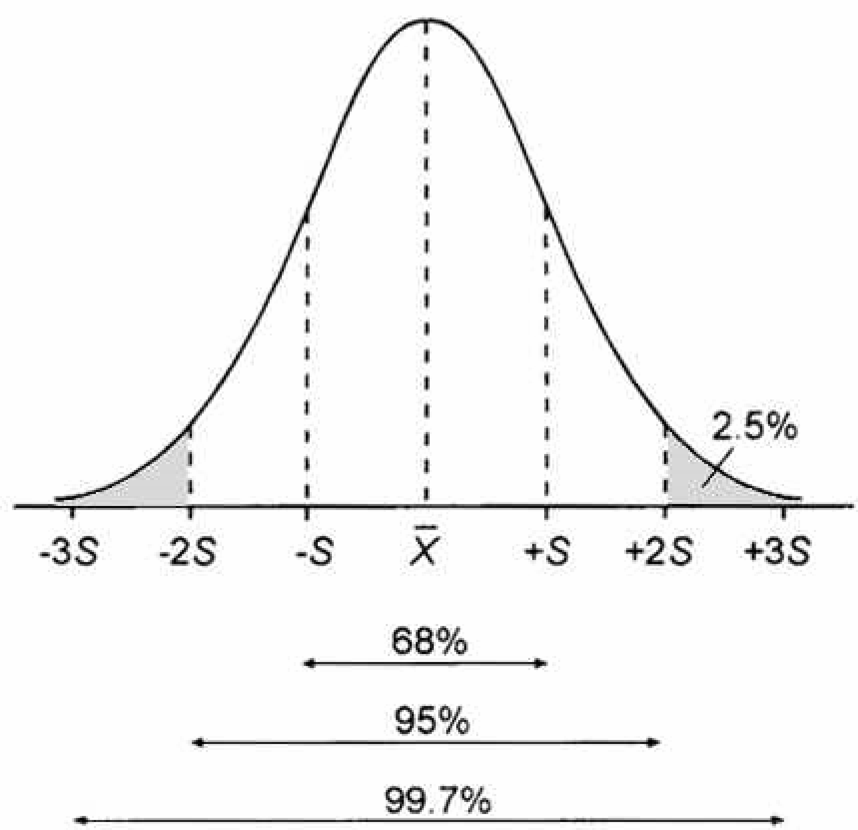
  - **2 S.D.** means that about **95%** of the data falls within $\bar{x} \pm 2 S.D.$

    EX: 75 ± 4.32; this means that 95% of the test scores fall between 70.68 and 79.32

  - **3 S.D.** means that about **99.7%** of the data falls within $\bar{x} \pm 3 S.D.$

    EX: 75 ± 6.48; this means that 99.7% of the test scores fall between 68.52 and 81.48

- **Assumptions:** standard variation assumes a **normal distribution** of the data. This means that most of the results should be close to the **mean** and the frequency of results decreases dramatically as one moves further away from the mean



Exercise: Find the **deviation** and **standard deviation** for the second set of test results [50, 60, 70, 80, 90, 100]