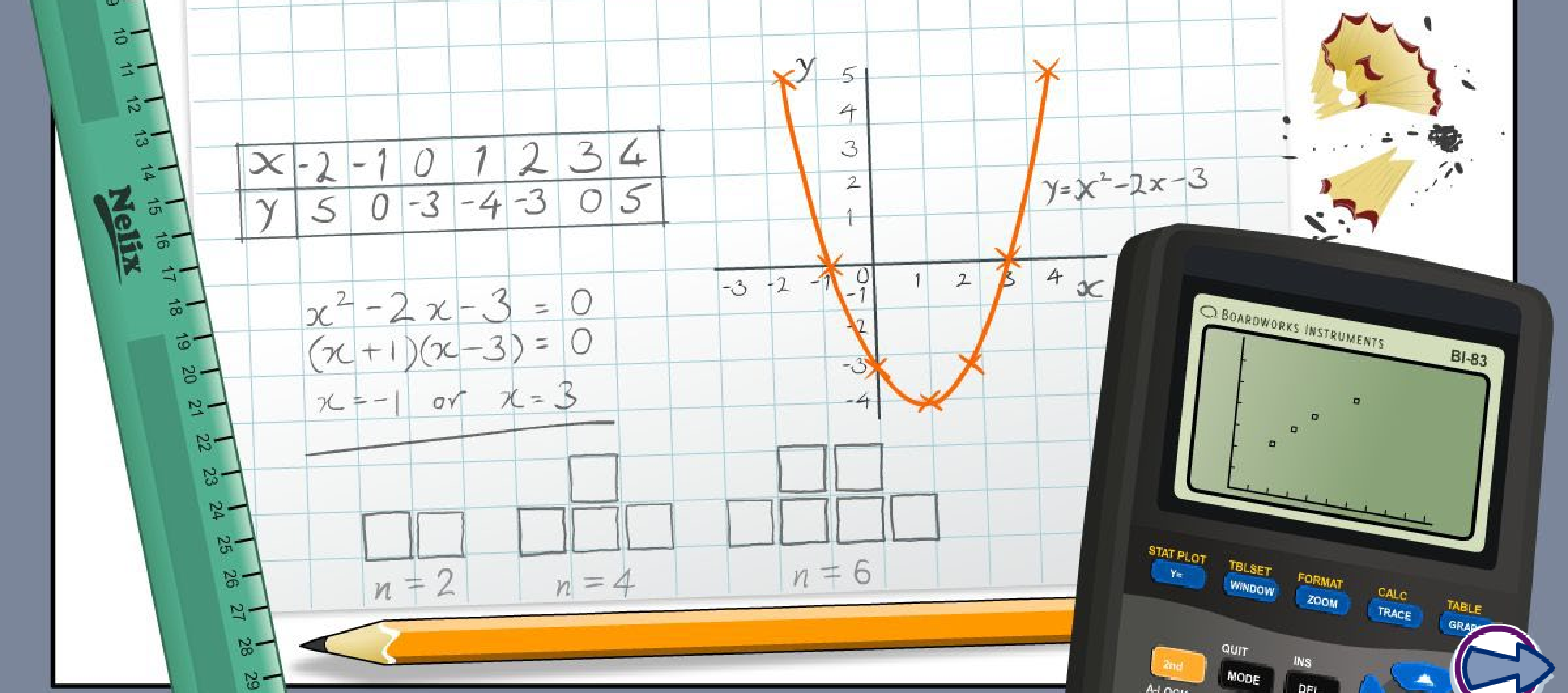
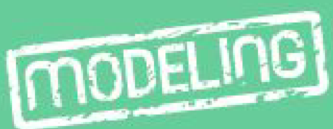
# Standard deviation



# Information

## Common core icons

This icon indicates a slide where the Standards for Mathematical Practice are being developed. Details of these are given in the Notes field.




Slides containing examples of mathematical modeling are marked with this stamp.

This icon indicates an opportunity for discussion or group work.

The **Standards for Mathematical Practice** outlined in the Common Core State Standards for Mathematics describe varieties of expertise that mathematics educators at all levels should seek to develop in their students.

They are:

1) **Make sense of problems and persevere in solving them.**
2) **Reason abstractly and quantitatively.**
3) **Construct viable arguments and critique the reasoning of others.**
4) **Model with mathematics.**
5) **Use appropriate tools strategically.**
6) **Attend to precision.**
7) **Look for and make use of structure.**
8) **Look for and express regularity in repeated reasoning.**

This icon indicates that the slide contains activities created in Flash. These activities are not editable.

This icon indicates teacher's notes in the Notes field.

# Consistency

The table shows the number of cars sold by the sales teams at two car dealerships in April 2012.

| dealership A | sales | dealership B | sales |
|---|---|---|---|
| Carla | 12 | Tony | 3 |
| Jacob | 8 | Lara | 10 |
| Jorge | 6 | Brandon | 10 |
| Aaron | 9 | Tess | 8 |
| Leticia | 13 | Amanda | 7 |

**How can we compare the consistency of the two sales teams? Find the mean sales at dealerships A and B and discuss how each salesperson compares to this mean.**

# What is standard deviation?

The **standard deviation** of a set of data is a measure of the distribution of the values. It indicates how far away the data values are from the **mean**.

A small standard deviation compared to the data values indicates that the data is packed closely around the mean.

A larger standard deviation results from a more widely distributed set of data.

Here are some symbols you will encounter when calculating standard deviation:

$x$ is a data value

$\bar{x}$ is used for the mean

$\sigma$ is used for the standard deviation

$\Sigma$ means “the sum of”

$n$ is the number of values in the data set

# The mean

The table shows the number of home runs scored by ten baseball players over the 2011 season.

| Name of player | Home runs | Name of player | Home runs |
|---|---|---|---|
| Mark Teixera | 39 | Johnny Damon | 24 |
| Alex Rodriguez | 30 | Jorge Posada | 22 |
| Nick Swisher | 29 | Derek Jeter | 18 |
| Hideki Matsui | 28 | Melky Cabrera | 13 |
| Robinson Cano | 25 | Jerry Hairston | 2 |

The mean number of home runs is:

$$\bar{x} = \frac{39 + 30 + 29 + ... + 2}{10} = \frac{230}{10} = \mathbf{23}$$

# Calculating standard deviation 1

The **standard deviation** (σ) is a measure of how far each data value is from the mean.

| value, $x$ | $x - \bar{x}$ |
|---|---|
| 39 | 16 |
| 30 | 7 |
| 29 | 6 |
| 28 | 5 |
| 25 | 2 |
| 24 | 1 |
| 22 | –1 |
| 18 | –5 |
| 13 | –10 |
| 2 | –21 |

**What do you think the first step is when finding the standard deviation?**

The first step is to subtract $\bar{x}$ from each data value.

Here, the mean is 23, so we subtract 23 from every piece of data.

**What do you notice about this column?**

All of these values add up to 0.

# Calculating standard deviation 2

| value, $x$ | $x - \bar{x}$ | $(x - \bar{x})^2$ |
|---|---|---|
| 39 | 16 | 256 |
| 30 | 7 | 49 |
| 29 | 6 | 36 |
| 28 | 5 | 25 |
| 25 | 2 | 4 |
| 24 | 1 | 1 |
| 22 | –1 | 1 |
| 18 | –5 | 25 |
| 13 | –10 | 100 |
| 2 | –21 | 441 |

**What can we do now to ensure that all these values are positive?**

We can square all of the values.

The second step is to find the values of $(x - \bar{x})^2$.

# Calculating standard deviation 3

The third step is to find the mean value of $(x-\bar{x})^2$.

| data value, $x$ | $x-\bar{x}$ | $(x-\bar{x})^2$ |
|---|---|---|
| 39 | 16 | 256 |
| 30 | 7 | 49 |
| 29 | 6 | 36 |
| 28 | 5 | 25 |
| 25 | 2 | 4 |
| 24 | 1 | 1 |
| 22 | –1 | 1 |
| 18 | –5 | 25 |
| 13 | –10 | 100 |
| 2 | –21 | 441 |
| | **Total:** | **938** |

The mean value of the last column is:

$$\frac{938}{10} = \mathbf{93.8}$$

This value is known as the **variance**: the mean of the squared deviations from the mean.

# Calculating standard deviation 4

| data value, $x$ | $x-\overline{x}$ | $(x-\overline{x})^2$ |
|---|---|---|
| 39 | 16 | 256 |
| 30 | 7 | 49 |
| 29 | 6 | 36 |
| 28 | 5 | 25 |
| 25 | 2 | 4 |
| 24 | 1 | 1 |
| 22 | –1 | 1 |
| 18 | –5 | 25 |
| 13 | –10 | 100 |
| 2 | –21 | 441 |
| | **Total:** | **938** |

In the process of finding the variance, we squared the values to make them all positive.

We therefore need to take the **square root** of the variance to find the standard deviation.

variance = 93.8

standard deviation $= \sqrt{93.8}$

$= \mathbf{9.69}$

(to nearest hundredth)

# What does it mean?

Here is the original table of data:

| name of player | home runs | name of player | home runs |
|---|---|---|---|
| Mark Teixera | 39 | Johnny Damon | 24 |
| Alex Rodriguez | 30 | Jorge Posada | 22 |
| Nick Swisher | 29 | Derek Jeter | 18 |
| Hideki Matsui | 28 | Melky Cabrera | 13 |
| Robinson Cano | 25 | Jerry Hairston | 2 |

The mean number of home runs is **23**. The standard deviation is **9.69**.

**Discuss what the standard deviation shows about the data. Another group of players had a standard deviation of 5.81. Which group had the better results? Justify your answer.**

# Dot plots

Match the means and standard deviations with the correct dot plots.

Press **start** to begin.

start

| | | | |
|---|---|---|---|
| Mean = 4.5<br>SD = 1.66 | Mean = 8.5<br>SD = 2.6 | Mean = 4.5<br>SD = 2.6 | Mean = 8.5<br>SD = 1.66 |

# General formulas 1

The **variance** of a set of data is given by the formula:

$$\text{variance} = \frac{\sum(x - \bar{x})^2}{n}$$

The **standard deviation** is the square root of the variance:

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

In practice, the following equivalent formulas are usually used for finding the mean and the variance:

$$\text{variance} = \frac{\sum x^2}{n} - \bar{x}^2$$

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

The key quantity in these two formulas is: $\sum x^2$.

This is the sum of the squares of each piece of data.

# Using the general formula

To use the formula, we need to find $\sum x^2$.

| data value $x$ | $x^2$ |
|---|---|
| 39 | 1521 |
| 30 | 900 |
| 29 | 841 |
| 28 | 784 |
| 25 | 625 |
| 24 | 576 |
| 22 | 484 |
| 18 | 324 |
| 13 | 169 |
| 2 | 4 |
| **TOTAL** | 6228 |

Substituting $\sum x^2 = 6228$ and $\bar{x} = 23$ into the formula for the variance gives:

$$\text{variance} = \frac{\sum x^2}{n} - \bar{x}^2$$

$$= \frac{6228}{10} - 23^2$$

$$= \mathbf{93.8}$$

$$\text{standard deviation} = \sqrt{93.8}$$

$$= \mathbf{9.69}$$

(to nearest hundredth)

# Practice

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 9 | 2 | 8 | 4 | 6 | 8 | 1 | 5 | 5 |

$\sum x =$ ______

$\sum x^2 =$ ______

Mean = $\frac{\_\_\_}{\_\_\_}$

Variance = $\frac{\_\_\_}{\_\_\_} -$ ______

Mean = ______

Variance = ______

SD = ______

next

# Using a calculator

You can use your graphing calculator to help you easily calculate the standard deviation of a data set.

Press **start** to see the steps (based on a TI-83/84), then press the **forward** arrow to work through them in detail.

start