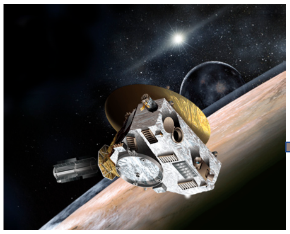


A Dusty Dilemma

Lesson Summary

Students will be introduced to the concepts of error analysis, including standard deviation. They will apply the knowledge to two sets of data from SDC.

Prior Knowledge and Skills

- Familiarity with squares, summation notation and square root function
- Knowledge of finding averages and means

AAAS Science Benchmarks

The Nature of Mathematics

Mathematical Inquiry

The Mathematical World

Reasoning

Habits of Mind

Computation and Estimation

NSES Science Standards

Science as Inquiry

Abilities Necessary to do Scientific Inquiry

Understandings About Scientific Inquiry

Teaching Time: 1 50-minute period(s)

Materials

Per student:

- 1 calculator
- Colored pencils (optional)

To share with class:

- 1 tape measure
- White board or chalk board
- 2 contrasting colors of marker or chalk for the board

Advanced Planning

Preparation Time: ~20 minutes

Gather materials and familiarize yourself with the SDC New Horizons website http://lasp.colorado.edu/sdc

LASP
LABORATORY FOR ATMOSPHERIC AND SPACE PHYSICS
UNIVERSITY OF COLORADO AT BOULDER

Educator Guide and Lesson Key

In this activity, we explore the concepts of averages (means), standard deviation from the mean, and error analysis. Students explore the concept of standard deviation from the mean before using the Student Dust Counter (SDC) data to determine the issues associated with taking data including error and noise. Questions are deliberately open-ended to encourage exploration.

Time: 1 50-minute period
Grade level: 8-10
Group size: 1-2

Materials

Per student
1 calculator
Colored pencils (optional)

To share with the class:
1 tape measure
White board or chalk board
Several dry erase markers or pieces of chalk in 2 contrasting colors (i.e. red and blue)

Prior Knowledge and Experience

- Familiarity with squares, summation notation, and square root function
- Knowledge of finding averages and means

Skills Used

- Interpreting data
- Graphing data
- Computing averages
- Computing standard deviation

Procedure

1. Choose a marker or chalk color to represent the boys in the class, and a contrasting color to represent the girls in the class.
2. Distribute the student handout.
3. Draw a table on the board for the height of the boys and girls as shown on student handout.
4. Draw 2 axes, one for a “Boy Plot” and another for a “Girl Plot” with height on the x-axis and “number of students” on the y-axis.

Label one "Boys" and the other, "Girls." Color-code the plots to match the color chosen for each gender.

5. In Part I, have each student write his or her height on the board under either the boy or girl column in inches, rounded to the nearest half inch. Have a boy representative and a girl representative plot the heights v. number of students. A scatter plot is preferable, but a bar chart could also be used. Afterward, students will compute the average height of each gender, the deviation from the mean for each measurement, and the average and standard deviation for the entire class, as outlined on the student handout. Draw a vertical line through the average value, and plus and minus the standard deviation for each plot. Ask students what they notice about the distribution of students on the plots. Do all students fall within the standard deviation?
6. In Part II, students will be given an SDC data set and will compute standard deviation based upon the data and will plot their results.

References:

Margaret A. McDowell, M.A., Fryar, C.D., Ogden, C.L., & Flegal, K.M. (2008, October 22). Anthropometric Reference Data for Children and Adults: United States, 2003–2006, National Health statistics report, 10. Retrieved from: **http://www.cdc.gov/nchs/data/nhsr/nhsr010.pdf**

Montgomery, D.C., & Runger, G.C. (1999). Applied Statistics and Probability for Engineers (2nd ed.), 1-9. New York: John Wiley & Sons, Inc.

For degrees of freedom description, the following website proved useful:

http://onlinestatbook.com/chapter8/df.html

Educator Lesson Key

Part I

I.I Include yourself in the data below.

Results will vary.

Number of girls in class today:

Number of boys in class today:

I.II

Record your height in the appropriate column on the board. From the board, record the height of the students in your class below.

Results will vary.

Height of Boys (inches)	Boy deviation from the mean (inches)	Height of Girls (inches)	Girl deviation from the mean (inches)

I.III

To find the average (or mean) height, write the equation below:

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

Where x is the height and n is the number of students.

MEAN GIRL HEIGHT (in class):

Answers will vary.

MEAN BOY HEIGHT (in class):

Answers will vary.

I.IV

a. How does your height *deviate* from the mean height? Subtract the mean height from your height to get the difference. This is the *deviation from the mean* value. (NOTE: If the value is negative, it indicates you are shorter than the class average, and if it is positive, you are taller):

Answers will vary.

b. Record this number in the appropriate column next to your height on the board, and record all of the class' values in your table.

I.V

What we want is the *standard deviation*. A standard deviation tells you how spread out the data is from the mean. (For example, if the mean height for the girls is 62 inches (5'2") the standard deviation might be something like 3 inches. That tells you that *most* of the girls are within 3 inches of the mean height).

a. You *could* just take the average of the deviations from the mean you recorded in the table, but what happens when you add up the values from that column? Explain why you get this result:

The answer will come out to be zero every time because the mean value is exactly in the center of all of the values. It is the definition of average (or mean). When you subtract the average value from the heights, you will always get an equal value above and below the average value. Accept a variety of reasonable answers.

b. We don't really care if the deviations are positive or negative. To fix this problem, we can square all of *deviations from the mean*, add those up, take the average of that, and then take the square root. That gives you the *standard deviation.*

The equation looks like this:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$

σ is the standard deviation,

n is the number of measurements,

x_i is each measurement recorded (in our case, each of the heights)

$\bar{x}$ is the mean value of the measurements

Compute the standard deviation using the equation above:

Answers will depend on the heights in the class. The answer should be reasonably small (within 5 inches) unless you have a large distribution of heights. The higher the value, the larger the class distribution is.

Standard deviation for boys:

Standard deviation for girls:

After the standard deviation has been computed, have two students (one male, one female) plot height on the x-axis and number of students on the y-axis for each gender. Have them draw vertical lines representing the average, and two dotted lines representing the standard deviation around the average (see "Student Height in Class" plot).

I.VI

You might be wondering why we divide by n-1, instead of just n. Aren't we just taking an average of all the *deviations from the mean*?

Try this: Compute the standard deviation for boys and then girls, but instead of using (n-1) in the denominator, just use n. What happens and why?

You will get a smaller standard deviation by using only n because you are dividing by a larger number in the denominator.

I.VII

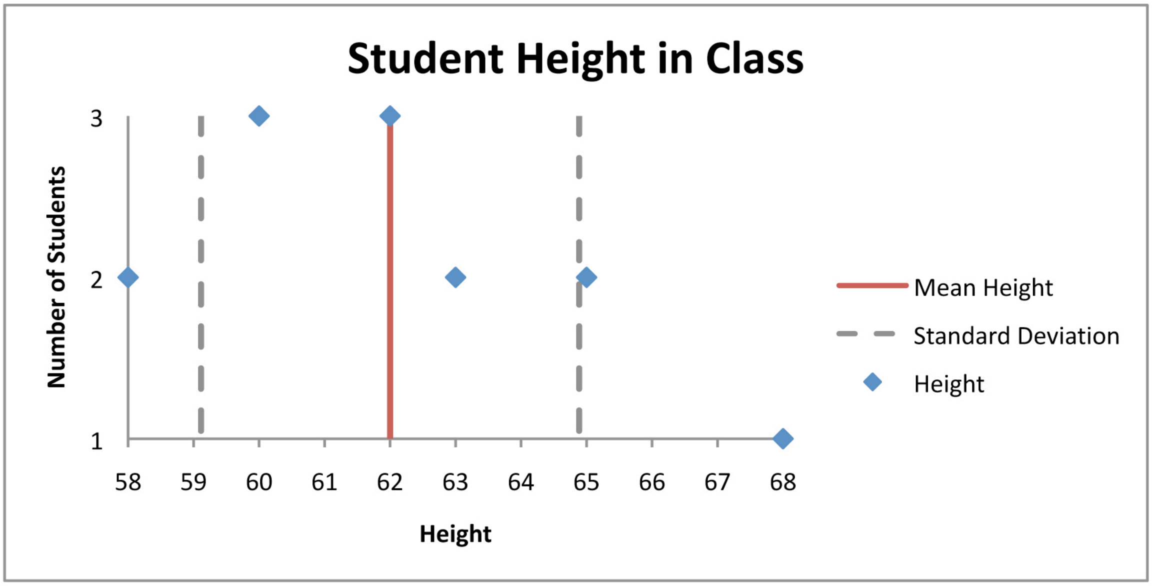
We are estimating the standard deviation using the assumption that your class represents a typical class. We've sampled your class and now we want to use your class to estimate the average height of all students your age. Do you know how many girls and boys your age are in your school and their heights? How about your city? What about your state? The whole country? The world? You probably don't know these numbers. This makes it more complicated to find an exact value for *standard deviation*.

Since we have estimated one thing in the equation, the mean value, we subtract the number of things estimated (remember, that's one thing) from the number of measurements to give us something called *degrees of freedom*. The idea is that if someone tall walked into the room, they would be more likely to fall within the standard deviation you've calculated. Using n-1 in the denominator instead of just n gives us a better idea of the true standard deviation for everybody everywhere.

From the plot on the board, what do you notice about the standard deviation lines? Does anyone fall above or below the lines? Explain the plot:

View the plot: "Student Height in Class"

This is an example of what a reasonable plot might look like. The red line shows the mean value for height, the gray dotted lines show the standard deviation around the mean value, and the blue diamonds show student heights. The x-axis is height, and the y-axis is number of students at that *height. The majority of the students should fall within the standard deviation from the mean (gray dotted lines). It's possible for some students to fall above or below the lines as in the "Student Height in Class" plot demonstrates. Although not covered in this assignment, approximately 68% of the class should fall within the standard deviation lines. Outliers are typically far above or below the average value. Be sensitive to the fact that some students may be embarrassed about their heights when discussing the topic. Since students are still growing, it is likely the difference between minimum and maximum height is large and this is just an expected result.*



Note: If you would like to compare the class' calculations with national data, the reference data that follows are nationally accepted values for height for students of different ages, courtesy of the CDC. Assume students are essentially the same age when comparing your class' values with the accepted values.

MALES:

Age	Sample size, n	Mean height (inches)	Standard Deviation
13	284	63.7	0.34
14	260	66.4	0.28
15	270	68.3	0.24
16	308	69.2	0.26
17	278	69.5	0.19
18	284	69.6	0.21
19	271	69.6	0.36

FEMALES:

Age	Sample size, n	Mean height (inches)	Standard Deviation
13	292	62.4	0.24
14	270	63.2	0.23
15	254	63.8	0.24
16	261	64.1	0.23
17	275	63.9	0.16
18	304	64.2	0.19
19	267	64.2	0.23

Part II

New Horizons and the Dusty Dilemma

II.I

New Horizons, while traveling on its path through space, collects dust that impacts the Student Dust Counter instrument. Every time a particle of dust hits one of SDC's 14 detectors, a small electrical charge is created, and it "counts" the dust. The thing is, sometimes it records a dust hit when no dust has impacted. A lot of different things could make the detector think it's recording a dust hit, when it's really not. Electrical impulses from the electronics onboard the spacecraft, for instance, might make the detector think it's looking at dust. This is called noise. Fortunately, we can remove the noise with some degree of *statistical certainty*.

The Student Dust Counter has fourteen channels. Twelve of the channels are science channels, collecting dust. Two of the channels, number 7 and 14, are not collecting dust. They are positioned so they point inward toward the spacecraft's interior, instead of outward exposed to space. Any measurement they make, then, is not dust at all but is noise. We can remove the noise from the other channels so we know how many counts are real dust hits and which are not.

Below are two tables, each representing the average number of counts of dust hits over a month for each channel.

April 2009

Channel	Hits
1	132
2	134
3	130
4	117
5	111
6	128
7	87
8	135
9	110
10	92
11	105
12	113
13	79
14	91

September 2009

Channel	Hits
1	56
2	31
3	44
4	40
5	45
6	34
7	34
8	33
9	47
10	37
11	39
12	41
13	25
14	23

Note: Discourage students from writing down a long string of decimal places. The number of significant figures should not be higher than the number of significant figures shown in the data tables above.

II.II

For the following calculations, round your answers:

a. Compute the average number of reference counts from channels 7 and 14, $\bar{x}_{\text{ref}}$.

Sum channels 7 and 14, and divide by 2, using equation:

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

Where n is equal to 2.

April 2009: *89*

September 2009: *28.5 (round to 29)*

b. Compute the average number of counts from all of the science channels, $\overline{x}_{sci}$.

Sum all science channels (all except 7 and 14), using equation:

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

Where n is equal to 12.

April 2009: *115.5 (round to 116)*

September 2009: *39.3 (round to 39)*

c. Compute the standard deviation for the reference channels, σ_{ref}.

Using equation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n - 1}}$$

σ is the standard deviation,
n is equal to 2
x_i is each measurement made by the reference channels
$\overline{x}$ is the mean value of the measurements from part a

April 2009: *2.83*

September 2009: *7.8*

d. Compute the standard deviation for the science channels, σ_{sci}.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$

σ is the standard deviation,

n is equal to 12

x_i is each measurement made by the science channels

$\bar{x}$ is the mean value of the measurements from part b

April 2009: *17.6*

September 2009: *8.2*

II.IV

a. How would you find the actual average number of dust hits?

To do this, subtract the average number of reference counts from the average number of science counts, using equation:

$$\bar{x}_{actual} = \bar{x}_{sci} - \bar{x}_{ref}$$

b. Compute the average number of dust hits, $\bar{x}_{actual}$.

April 2009: 26.5
September 2009: 11

II.V

Now, we want just the standard deviation for the actual dust hits. We have a fair amount of uncertainty in the measurements in each of the science channels and in the reference channel. If we just subtracted the standard deviation of the reference channels from the other channels, we would get a smaller value for uncertainty in the average number of dust hits when really we should have more uncertainty in the final value. The total actual standard deviation for the dust hits, then, is found using this equation:

$$\sigma_{actual} = \sqrt{\sigma_{sci}^2 + \sigma_{ref}^2}$$

a. Compute σ_{actual} for the two data sets:

April 2009: 17.8

September 2009: 11

b. Fill in the table below:

Month	Actual average number of dust hits	Actual standard deviation
April	80.8	53.3
September	11	11

c. Explain what the standard deviation for the above data is telling us:
In September, the standard deviation is as large as the number of dust hits recorded giving 11± 11. This means that the error bars include zero, so it is possible that none of the dust recorded by the detector is real dust at all. In April, the standard deviation is pretty high. The number of dust hits is about 81 ± 53, which indicates that there is a high value for error, but that real dust was actually recorded by the detectors. Accept a variety of reasonable responses. Discuss with students the idea that probability is an essential part of science experiments. Often, we have to improve equipment or experimental parameters to narrow down a real value, but error exists even in the most basic calculation or with the best tools. Take a ruler, for example; when you make a measurement using a ruler, you are relying upon the idea that the ruler gives an accurate and precise measurement. This is not the case, however. When the ruler was created, there was a probability for error in the placement of each marker, each of the lines has it's own width that factors into the measurement, and very small differences in length cannot be recorded by a ruler. If a set of calipers is used, instead, your degree of certainty in the measurements goes up, but it is also not without error.

NAME

Part I

I.I Include yourself in the data below.

Number of girls in class today:

Number of boys in class today:

I.II

Record your height in the appropriate column on the board. From the board, record the height of the students in your class below.

Height of Boys (inches)	Boy deviation from the mean (inches)	Height of Girls (inches)	Girl deviation from the mean (inches)

I.III

To find the average (or mean) height, write the equation below:

MEAN GIRL HEIGHT (in class):

MEAN BOY HEIGHT (in class):

I.IV

a. How does your height *deviate* from the mean height? Subtract the mean height from your height to get the difference. This is the *deviation from the mean* value. (NOTE: If the value is negative, it indicates you are shorter than the class average, and if it is positive, you are taller):

b. Record this number in the appropriate column next to your height on the board, and record all of the class' values in your table.

I.V

What we want is the *standard deviation*. A standard deviation tells you how spread out the data is from the mean. (For example, if the mean height for the girls is 62 inches (5'2") the standard deviation might be something like 3 inches. That tells you that *most* of the girls are within 3 inches of the mean height).

a. You *could* just take the average of the deviations from the mean you recorded in the table, but what happens when you add up the values from that column? Explain why you get this result:

b. We don't really care if the deviations are positive or negative. To fix this problem, we can square all of *deviations from the mean*, add those up, take the average of that, and then take the square root. That gives you the *standard deviation*.

The equation looks like this:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$

σ is the standard deviation,

n is the number of measurements

x_i is each measurement recorded (in our case, each of the heights)

$\bar{x}$ is the mean value of the measurements

Compute the standard deviation using the equation above:

Standard deviation for boys:

Standard deviation for girls:

I.VI

You might be wondering why we divide by n-1, instead of just n. Aren't we just taking an average of all the *deviations from the mean*?

Try this: Compute the standard deviation for boys and then girls, but instead of using (n-1) in the denominator, just use n. What happens and why?

I.VII

We are estimating the standard deviation using the assumption that your class represents a typical class. We've sampled your class and now we want to use your class to estimate the average height of all students your age. Do you know how many girls and boys your age are in your school and their heights? How about your city? What about your state? The whole country? The world? You probably don't know these numbers. This makes it more complicated to find an exact value for *standard deviation*.

Since we have estimated one thing in the equation, the mean value, we subtract the number of things estimated (remember, that's one thing) from the number of measurements to give us something called *degrees of freedom*. The idea is that if someone tall walked into the room, they would be more likely to fall within the standard deviation you've calculated. Using n-1 in the denominator instead of just n gives us a better idea of the true standard deviation for everybody everywhere.

From the plot on the board, what do you notice about the standard deviation lines? Does anyone fall above or below the lines? Explain the plot:

Part II

New Horizons and the Dusty Dilemma

II.I

New Horizons, while traveling on its path through space, collects dust that impacts the Student Dust Counter instrument. Every time a particle of dust hits one of SDC's 14 detectors, a small electrical charge is created, and it "counts" the dust. The thing is, sometimes it records a dust hit when no dust has impacted. A lot of different things could make the detector think it's recording a dust hit, when it's really not. Electrical impulses from the electronics onboard the spacecraft, for instance, might make the detector think it's looking at dust. This is called noise. Fortunately, we can remove the noise with some degree of *statistical certainty*.

The Student Dust Counter has fourteen channels. Twelve of the channels are science channels, collecting dust. Two of the channels, number 7 and 14, are not collecting dust. They are positioned so they point inward toward the spacecraft's interior, instead of outward exposed to space. Any measurement they make, then, is not dust at all but is noise. We can remove the noise from the other channels so we know how many counts are real dust hits and which are not.

Below are two tables, each representing the average number of counts of dust hits over a month for each channel.

April 2009

Channel	Hits
1	397
2	403
3	389
4	351
5	332
6	385
7	260
8	404
9	331
10	277
11	314
12	340
13	238
14	272

September 2009

Channel	Hits
1	56
2	31
3	44
4	40
5	45
6	34
7	34
8	33
9	47
10	37
11	39
12	41
13	25
14	23

II.II
For the following calculations, round your answers:

a. Compute the average number of reference counts from channels 7 and 14, $\bar{X}_{\text{ref}}$.

April 2009:

September 2009:

b. Compute the average number of counts from all of the science channels, $\bar{X}_{\text{sci}}$.

April 2009:

September 2009:

c. Compute the standard deviation for the reference channels, σ_{ref}.

April 2009:

September 2009:

d. Compute the standard deviation for the science channels, σ_{sci}.

April 2009:

September 2009:

II.IV
a. How would you find the actual average number of dust hits?

b. Compute the average number of dust hits, $\bar{X}_{\text{actual}}$.

April 2009:

September 2009:

II.V

Now, we want just the standard deviation for the actual dust hits. We have a fair amount of uncertainty in the measurements in each of the channels and in the reference channel. If we just subtracted the standard deviation of the reference channels from the other channels, we would get a smaller value for uncertainty in the average number of dust hits when really we should have more uncertainty in the final value. The total actual standard deviation for the dust hits, then, is found using this equation:

$$\sigma_{\text{actual}} = \sqrt{\sigma_{\text{sci}}^2 + \sigma_{\text{ref}}^2}$$

a. Compute σ_{actual} for the two data sets:

April 2009:

September 2009:

b. Fill in the table below:

Month	Actual average number of dust hits	Actual standard deviation
April		
September		

c. Explain what the standard deviation for the above data is telling us: