



A Dusty Dilemma

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.1

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, $\sigma_{\text{ref.}}$

April 2009:

September 2009:

- d. Compute the standard deviation for the science channels, $\sigma_{\text{sci.}}$

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: