
THE STANDARD DEVIATION

“I may venture to say that *his* observations have stretched much farther...”
-- *Sense and Sensibility*

2.1 INTRODUCTION

It is sometimes said that “physics is an exact science.” But physics is based on experiment, and experiment on measurement, and measurement is inexact by its very nature. Imperfect (or simply just finite) instruments, necessary approximations, and inevitable background noise make all measured values impossible to pin down exactly. Even so, clever experiments that coax even a very uncertain but crucial measurement from a sea of noise can lead to great advances in physics. A recent article in a science magazine trumpeted the fact that, with the help of data from the Hubble Space Telescope, the various scientific groups measuring the Hubble Constant (a quantity whose value determines the age and fate of the universe) were now getting results that agreed to within roughly 10% of each other. This was considered great news, partly because it was a dramatic improvement over the situation just a few years ago (when measurements disagreed by more than a factor of two), but partly because the age of the universe computed from this quantity was finally settling down toward values that were consistent with other physical estimates of the age of the universe, soothing physicists’ worries that whole areas of physics might be poorly understood.

The point is that, while the greatest precision is always desirable, even inexact measurements can lead to important scientific progress. But a crucial part of making good scientific use of inexact results is being able to *quantify* how inexact (the technical word is **uncertain**) a measurement is. We can draw scientific conclusions from an inexact value only if we know something about how sharp or fuzzy the value is.

One of the major goals of this laboratory program is to teach you how to extract the maximum possible scientific meaning from uncertain measurement results. The next chapter of this reference manual will begin discussing the meaning of the concept of uncertainty in the context of measured quantities. The purpose of this chapter is to lay some mathematical foundations for that discussion by exploring ways that we might quantify the spread in a set of values that represent imperfect measurements of the same quantity.

2.2 POSSIBLE MEASURES OF SPREAD IN A DATA SET

Imagine that N teams of scientists measure a physical quantity, and each group ends up with a value that is somewhat different from the values obtained by other groups. This process gives us a data set consisting of N values, each of which represents an imperfect estimate of the “true value” of that quantity. As we will see in the next chapter, it is useful when discussing the physical meaning of these results to be able to quantify the “spread” in these measurement values. How might we quantitatively express the “spread” in a data set like this in a meaningful and useful way?

The simplest way to express the spread in a data set is to use its **range** R , which in this context we define to be the difference between the highest and lowest values in the set. Clearly, the more “spread out” a set of data values is, the larger its range is likely to be.

A potential problem with the range as a measure of the spread in a data set is that its value is determined entirely by the two most extreme values in the set (all the other values are ignored!) The extreme value may be the result of improbable flukes in the measurement process (or even outright

errors), and are not representative of the whole by their very nature. One might imagine one data set where the values are mostly very closely clustered around some central value with only a handful of wildly different values at the extremes, and another data set having the same range whose values are fairly evenly spread out between the extremes. Would we really want to quantify the “spread” in these two data sets by the same number?

A more meaningful way to quantify the spread is to compute the **average deviation** of the data points from the data set’s **mean** (= average) value. If we use the symbols x_1, x_2, x_3, \dots to represent the first, the second, the third, etc. measurement values in our set of N values, and the symbol \bar{x} to stand for the mean of those values, then the average deviation is defined to be:

$$\text{average deviation} \equiv \frac{1}{N} \left[|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \dots + |x_N - \bar{x}| \right] \quad (2.1)$$

You can see that this definition gives you exactly what the name implies, the *average* (summed over the entire data set) of the *deviations* of the measurements (which for a given measurement is the absolute value of how far that measurement is from the mean). Note that the absolute value is essential in this expression: since generally a given measurement is as likely to be above the mean as below it, if we did not take the absolute value of each difference, they would sum to zero (in fact the definition of the mean implies that if we remove the absolute value symbols in equation 2.1, the sum would be *exactly* zero, as you can check).

The average deviation, in contrast to the range, nicely takes each measurement value equally into account in its value, and thus is probably a better representation of the spread in a data set. It also has an easily understood meaning. However, it turns out that we often want to do some calculus with the quantity we use to characterize the spread in data. The absolute value presents a minor complication in doing calculus with the average deviation, because the derivative of $|x|$ is undefined when $x = 0$.

The **standard deviation** does not suffer from this problem. If we use the same symbols that we used in equation 2.1, we can write the standard deviation as follows:

$$\text{standard deviation} \equiv s \equiv \sqrt{\frac{1}{N-1} \left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_N - \bar{x})^2 \right]} \quad (2.2)$$

If we ignore for a moment the fact that we are dividing by $N-1$ instead of N , the standard deviation is thus the *square root of the average squared deviation*. Dividing by $N-1$ instead of N is important for deep mathematical reasons beyond our level here, but note that for a data set consisting of a single measurement, the average deviation states that the “spread” in this data set is *zero* (suggesting that we know the measurement value perfectly), whereas the standard deviation states that the spread is *undefined* ($0 \div 0$), which more meaningfully suggests that we know nothing about the spread in a data set when that set only has one value.

Equation 2.2, though it superficially looks more complicated than that for the average deviation, is actually simpler in a number of respects. Doing calculus with this expression is easier because squares and square roots are easier to deal with mathematically than the absolute value. Since many calculators are specially set up to calculate the standard deviation with a few button presses, it is generally much easier to calculate than the average deviation, and is really no more difficult to calculate even if one is reduced to using equations 2.1 and 2.2 directly. However the *main* reason that the scientific community typically uses the standard deviation (rather than the range or the average deviation) to characterize the spread in the values of a data set is that it has an important property that the range and average deviation do not have.

Imagine that we model each of the N measurements in our data set as being what happens when we add tiny random errors from fairly large number of sources to the measured quantity's "true value" (whatever that might be). This "random error" measurement model is only a model of the measurement process (and a pretty simplified one at that), but it does seem to be useful and reasonably accurate in many cases. When this model does adequately describe a measurement process, then we find that if we plot the probability of getting a certain measurement value versus that value, we get the famous "bell-shaped curve" (which indicates that measurement values close to the quantity's "true value" are common and extreme values are rare). A wide variety of measurement process tend to yield data sets that are distributed this way.

So here is the payoff. If the "random error" measurement model is a good model for the measurement process giving rise to given data set, *then* it can be shown mathematically that the value one gets by computing the standard deviation for this data set is fairly independent of N . This means that in the context of this "random error" measurement model, the standard deviation better isolates the effect of the *measurement process itself* on the spread in the data from effects that result from including more or fewer data points in the data set. In contrast, the *range* of a data set typically *increases* as N increases (because the more measurements we take, the more likely we are, by chance, to encounter some really extreme measurement results), while the *average deviation* tends (more subtly and for more subtle reasons) to *decrease* with increasing N . It is this special relationship between the standard deviation and the most common and useful mathematical model for the measurement process that makes the standard deviation the most widely accepted measure of the spread of the values in a data set.

2.3 CALCULATING THE STANDARD DEVIATION

One can fairly easily calculate the standard deviation of a list of measurement values directly from equation 2.2 by going through the following steps:

1. Write the values in a vertical column.
2. Sum the values and divide by the number of measurements N to find the mean.
3. Subtract this mean from each value and square the result. Write each "squared deviation" to the right of the corresponding measurement value.
4. Sum the squared deviations in this new column and divide by $N-1$.
5. Take the square root of the result to get the standard deviation.

Almost any scientific calculator (which we recommend you have for this course) will provide an easier way for you calculate the standard deviation. In general, the process looks like this:

1. Initialize the calculation somehow (varies from calculator to calculator).
2. Type in the first number.
3. Press a key that is usually marked $\Sigma+$ or $M+$.
4. Repeat steps 2 and 3 for each of the measurement values (after you press $\Sigma+$ each time, the calculator typically displays the number of values that you have entered so far).
5. When all values have been entered, calculate the standard deviation by pressing one of the following: s , s_x , σ , σ_{n-1} (if the choice is between that and σ_n), σ_x or $\sigma_{x_{n-1}}$ (instead of σ_y or $\sigma_{y_{n-1}}$). The choice, unhappily, varies from calculator to calculator.

For example, on my HP-32S calculator, one initializes the calculation using the CLEAR button and selecting "clear all" from the choices presented. To get the final result, I have to push the STAT button and then select " s " and then " s_x " from the choices presented. On a TI-36X calculator, one initializes the calculation by selecting the STAT 1 function (to indicate we are going to do statistics of one variable), and one gets the final result by pressing the $\sigma_{x_{n-1}}$ button.

On a TI-82 or TI-83 calculator, select the STAT key, used the arrow keys to get to EDIT if you aren't there already, and enter your data in one of the data lists (L1, L2, L3, ...). Then press the STAT key again, use the right-pointing arrow key to choose CALC, the down-arrow key (if necessary) to get to 1-Var Stats, and press ENTER to get the mean and standard deviation, as well as some sums used to calculate their values.

You should look in the instruction booklet that came with your calculator to determine how to calculate the standard deviation on your calculator. (If you don't have your booklet, do some experiments, ask a friend, or check your calculator manufacturer's web site. If you come a bit early to lab, your lab instructor or a lab assistant *may* be able to help you.) To test that you are actually calculating the standard deviation as defined by equation 2.2, note that the standard deviation of the data set 2, 4, 6 should be 2.

EXERCISES

Exercise 2.1

Compute the standard deviation of the following data set using the direct method (the first method outlined in section 2.3). Show your work in the space below (be sure to keep track of units!). Then figure out the particular short-cut approach that works on your calculator and use it to check your work.

0.56 s

0.52 s

0.59 s

0.48 s

0.51 s

Exercise 2.2

Show your interviewer that you can calculate the standard deviation of the data set 2, 4, 6 using whatever short-cut approach works on your calculator.