

## Error.

The most consistently misunderstood and mis-handled concept in the chemistry lab (or any other subject that uses measured numerical data) is the idea of error. This is partially because when most people hear “error”, they think “wrong”. In this case, however, error simply means the variability or uncertainty in a reported value based upon measured values. Experimental error is not a bad thing, it’s an inevitable part of measured data and as such must be reported if you intend your audience to believe that your data is valid. There are a number of ways to treat error, let’s start with a very simple example so we can see how the error comes out of (and affects) the data.

You have gone fishing and you caught 4 walleye. Their lengths are listed in the following table.

Fish	Length
Walleye #1	11.5”
Walleye #2	28.0”
Walleye #3	21.5”
Walleye #4	18.0”

What was the average length of walleye you caught? OK, that’s almost self explanatory, you take the average.

$$\text{Average length} = \frac{\text{sum of lengths}}{\text{number of fish}} = \frac{11.5''+28.0''+21.5''+18.0''}{4} = 19.75''$$

Great, what does it mean? {Ha, ha, there’s a math joke rolled into that question...} None of the fish you caught were 19.75” long, so there must be some variability in that data. If you tell a friend that you caught four walleye with an average length of 19.75”, that might mean that they were all keeper size and you’re going to have a nice dinner, or it might mean that they were all in the protected slot and had to be released. How can you more accurately report your day’s catch? Here comes the “reported numerical error”. The biggest fish was 28.0” and the smallest was 11.5”, so the range of sizes was (28.0”-11.5”=16.5”). The error (or variability) in the measurement is half the range, or 8.25”. Error should always be reported as one digit (unless the first digit is “1”, more on that later...), so the error should be rounded to 8”. Now, what about the average? If we’re saying that there’s error in the ones place, it really doesn’t make any sense to report the average to 2 decimals, so the reported average should be rounded to the same position as the error. SO, finally, you can report the average length of fish you caught as 20”±8”.

*A couple notes:*

1. Using “range / 2” for error only works reliably when your data is fairly evenly distributed. If the four fish were 15”, 17”, 14” and 28” and you treated the error the same way, you’d report the average length as 18” ±7”. Looking at the data, that reported value doesn’t really give an accurate picture of the observed lengths. In this case, you could probably assume that the 28” measurement was an outlier and disregard it in your analysis. (To be totally legitimate, you’d probably want to do a Q-test to see if that data point could be thrown out, feel free to look up Q-test if you’re interested...)
2. So what about those 15”, 17” and 14” fish? Their average is 15.33”, their range is 3, so the error is  $3/2 = 1.5$ ”. If we round that to a single digit, it rounds to 2, but that’s a problem. Why? Rounding 1.5 to 2 means we’re willing to accept 0.5 units of rounding error in our error. 0.5 is 33% of 1.5, so in a relative sense, we’re introducing a LOT of error by rounding. For that reason, it’s usually acceptable to keep two digits of error when the first digit is “1”, so we could report this as 15.3” ±1.5”.

### ***Absolute Error vs. Relative Error:***

This is more definitional, but depending upon what we’re looking at sometimes it makes more sense (and is more informative) to talk about absolute error and sometimes it’s better to use relative error. Absolute error is error that has the same units of measurement as the quantity we’re reporting. The errors above are absolute errors, “20 inches plus or minus 8 inches”. Relative errors are essentially like

percents, or maybe more correctly, fractions. (Think about them like the mol fraction or mass fraction concentration units.) Let's say I owe you \$750 and I pay you \$745. You're maybe not happy, but \$5 on a \$750 debt isn't that big of a deal. What if I owe you \$6 and I pay you \$1? The *absolute* error in my repayment is the same, \$5, but the *relative* error in my repayment is now 83% instead of 0.7%. That's a big difference.

### ***How does this relate to data you collect in lab?***

Many of you have asked "Can I take the average right away, or do I have to do the whole calculation for ALL of these runs and then take the average?" The answer is (as a number of you have heard this week), you can do it either way as long as you keep track of error. If you take the average *first*, then you have to include error *first* and you have to carry that error all the way through your calculation using a set of rules and guidelines usually called "propagation of error". When propagating error, sometimes you use absolute errors, sometimes you use relative errors, the rules change depending upon whether you're multiplying/dividing or adding/subtracting, and don't even get me started on what happens if you're doing logarithms and square roots and trigonometric functions. If you'd like to have some fun, check out [http://en.wikipedia.org/wiki/Propagation\\_of\\_uncertainty](http://en.wikipedia.org/wiki/Propagation_of_uncertainty). {Warning: If you're not a big fan of math, your head might explode.} If you'd like a slightly more complete and readable explanation of propagation of error, you might try

<http://www.physics.uc.edu/~bortner/labs/Physics%20Lab%20web%20site/Appendix%202/Appendix%202%20Error%20Propagation%20htm.htm>

I had to shrink it that much to fit it on one line, and it's from a physics guy so it's still got the math in it but it's pretty readable. Is this discouraging you from taking the average first? Hmm...

If you wait until the end to take an average, then you can (usually) treat the error in your measurements the same way we did in the fish example, and if you're using MSExcel or some other spreadsheet to do your calculations then it's no more difficult to calculate 1 line or 1000 lines of data, so waiting until the end to calculate an average and the error associated with that average sounds a little easier to me than all this propagation business...

### ***Can't I just calculate the standard deviation?***

Sometimes, but there are a few cautions here as well. First, if you don't think about what the standard deviation means, then the number you calculate will be meaningless. Is standard deviation equal to error? Is it an absolute error or a relative error? Second, standard deviation isn't always a very good way to think about error if you have a small sample size. Since most Gen Chem lab students only want to do 2 repeat runs (OK, maybe 3 if it's an easy experiment...), using standard deviation can be a little misleading. If you're looking at a set of dozens or hundreds or millions of individual measurements, standard deviation is great, but standard deviation becomes less and less reliable as sample size decreases, especially when sample size decreases to low single digits.

### ***The Bottom Line (for now...)***

This should get you through your next set of calculations, we'll probably revisit error down the road when other treatments are more appropriate. There are books upon books upon books FILLED with discussions of error (or uncertainty), and the field of statistics is almost all about error and uncertainty. At this point, the most important thing is that you try your best to include legitimate numerical error in any number you report. Like everything else, it takes practice, so try to calculate error whenever an appropriate data set presents itself. You and your friend each buy a bag of M&Ms? How many blue M&Ms are there in an "average" bag of M&Ms, including error? Going out to eat? What's the average price of an entrée on the menu, including error? A bunch of you are bored, short on cash and staring at your shoes? What is the average number of eyelets on all of your shoes, including error? Still bored and cashless? On average, how long can you all hold your breath, including error? What's the average number of pages in your textbooks this semester, including error? What's the average number of Facebook friends you all have, including error? It's everywhere, you just have to calculate it...