Numbers and Data Analysis

With thanks to George Goth, Skyline College for portions of this material.

Significant figures

Significant figures (sig figs) are only the first approximation to uncertainty and what is called ‘error analysis’. But what are sig figs? It depends on the circumstances. If your talking about a homework problem for instance, then it’s the number of digits given importance in a figure, usually two to three digits. If you are talking about data, experimental results, a theoretical or ‘known’ value (e.g. gravity) then it is the digits we are certain of plus one we are uncertain of.

Here are some rules for counting sig figs:

Zeros within a value are always counted.

Example: 4032 and 50.03 both have four sig fig.

Zeros setting the decimal point are not significant.

Example: All of these have 2 sig figs 52,000,000 3.7×10^9 0.00074 2.3×10^{-15}

Zeros not setting the decimal point are significant.

Example: 0.00025 has five sig figs.

Rounding numbers

Now I know that you paid a lot for that fancy scientific calculator that gives you something like ten digits or so, and you think “Hey, I paid for them, so I should use them!” well not so. Think of it this way: I have a friend who has a nice watch that is accurate to one second in ten thousand years, but when I ask him the time he tells me ‘About a quarter to twelve.’ Most times you only need two to four sig figs. Round your results to what is appropriate. How? look at the first number you are going eliminate. If it is five or greater increase the next number by one (rounding up) and if it is four or less leave that last number alone (rounding off). Here is an example:

rounding up 12.3456789 $\Rightarrow$ 12.35

rounding off 98.7654321 $\Rightarrow$ 98.77

Calculations and rounding

You have to be careful when rounding numbers. If you round too soon, the rounding will affect your results. An Example:

You have two measured values, 12.34 and 2.34. If you round, then multiply, then round your results to two sig figs you get very different outcomes.

Rounded to	Calculation		Result
Left alone	12.34 x 3.14 = 38.7476	⇒	39
One decimal place	12.3 x 3.1 = 38.1300	⇒	38
Whole number	12 x 3 = 36.0000	⇒	36

But when do you start rounding numbers? The simple answer is, as your last step. But you need to know what is significant during an intermediate step. Here are some examples of carrying sig figs through calculations.

For addition and subtraction. Your result has sig figs only to the decimal place that both the original numbers had sig figs.

12.35 cm	4 sig figs	Decimal to the hundredths
- 3.1 cm	2 sig figs	Decimal to the tenths

9.25 cm	calculator answer	

The value with the least number of decimal places was 3.1 cm, with a decimal place to the tenths, so the result is only be expressed to the tenths as well.

9.3 cm   Remember, fives round up.

Now for multiplication and division. The result has same number of sig figs as the factor with least number of sig figs.

	12.34 cm x	3.14 cm = 38.7476 cm	calculator answer
Sig figs	4	3	

So this result is expressed to 3 sig figs, which is 38.7 cm

Reporting values

First you report all numbers you are certain of and first you are uncertain of:

12.35 cm
c c c u

Where the uncertain number is ± 1

12.35 cm could be as low as 12.34 cm and as high as 12.36 cm

This is called the range of the value.

When you report a value, you first state the nominal value then the uncertainty. Using the previous example:

12.35 ± 0.01 cm
Nominal Uncertainty

Uncertainties

At this point you must be wondering were the uncertainties come from and if all values have an uncertainty. There are three types of values:

1) Constants and coefficients. These have no uncertainty associated with them. Examples are π, the natural log e, $\frac{1}{2}$ (e.g. $\frac{1}{2}\mathrm{m}\vec{\mathrm{v}}^2$), etc.

2) Measured values. These have an uncertainty that is associated with the device and method of taking the measurement (called systematic errors). Example:

Measuring the length 12.35 ± 0.01 cm with a centimeter rule. You are certain of the measurement of 12 cm, and 3 mm. You believe the true value of the last digit to lie about halfway between 3 mm and 4 mm and are confident in that give-or-take one unit (i.e. one tenth of a mm).

Measuring the mass of an object as 503 ± 0.5 gm on a beam balance. You are certain of the mass measurement 503 gm give-or-take half a gram.

3) Populations of measured values. These have an uncertainty that is associated with variations in the measurement (random errors). An example is you drop a ½ kg mass 100 times, measuring the final velocity each time. You now have a population of 100 data points measuring the velocity with a small (let's hope!) variation between the different measured values.

So the first case is easy to deal with, no uncertainty, the second and third need some attention.

Absolute and relative (percentage) uncertainties

You measure one mass with a precision of ± 1 mg; you measure another mass with a precision of ± 1 kg. Which measurement is more precise value?

It is tempting to say ± 1 mg is the more precise value, it is after all smaller, but that is not necessarily true that it is more precise. Suppose the ± 1 mg measurement was of an ant with a mass of 3 mg while the ± 1 kg measurement was of an elephant with a mass of 3000 kg.

Now ± 1 mg is 33% of 3 mg while ± 1 kg is 0.033 % of 3000 kg. So which is more precise now? It matters how big the uncertainty is compared to what you are measuring.

This illustrates the two types of uncertainty. Absolute (± 1 mg) and relative (or percentage) uncertainty. Absolute uncertainties (au)have units. Examples:

12.35 ± 0.01 cm or 503 ± 0.5 gm

Relative uncertainties (ru) have no units and are usually, though not always, shown as a percent (%).

ru = au/nominal value

percent uncertainty = ru x 100%

Example:

503 ± 0.5 gm

ru = au/ nominal value = 0.5/503 = 0.001
percent uncertainty = ru x 100 = 0.001 x 100 = 0.1 %

Combining uncertainties

When we combine values the uncertainties in the values have to be combined as well. This means that uncertainties propagate or grow. We use the following two rules.

1) Addition and subtraction. Add the absolute uncertainties. Examples:

Addition

add 12±2 gm to 97±2 gm

$$\begin{array}{r} 97 \pm 2 \\ \underline{+\ 12 \pm 2} \\ 109 \pm 4 \text{ gm} \end{array}$$

Subtraction

subtract 12±2 gm from 97±2 gm

$$\begin{array}{l} 97 \pm 2 \\ \underline{-\ 12 \pm 2} \\ 85 \pm 4 \text{ gm} \end{array}$$

2) Multiplication and division. Use the relative or percent uncertainty to find the combined uncertainty. Example

A desk top is measured to be 97 cm by 12 cm. If the absolute error on each is 2 cm, what is the area and the absolute, relative and percentage uncertainties?

Area = L W = 97 x 12 = 1164 cm^2

Now to find the uncertainties we must calculate the relative uncertainty of the length and width.

ru_{length} = 2 / 97 = 0.0206 (carrying a few extra sig figs for the moment)
ru_{width} = 2 / 12 = 0.1666
ru_{area} = ru_{length} + ru_{width} = 0.0206 + 0.1666 = 0.1873
au_{area} = Area x ru_{area} = 1164 x 0.1873 = 218 cm^2

Generally the absolute uncertainty is expressed to one sig fig (sometimes two sig figs) and the calculated value is expressed to the same decimal place as the absolute uncertainty

218 cm^2 = 200 cm^2 (1 sig fig) in the hundreds
or = 220 cm^2 (2 sig figs) in the tens

To the nearest hundred 1164 is 1.2×10^3 cm^2
with an uncertainty to one sig fig shown as $1.2 \pm 0.2 \times 10^3$ cm^2

or

To the nearest ten 1164 is 1.16×10^3 cm^2
with an uncertainty to two sig figs shown as $1.16 \pm 0.22 \times 10^3$ cm^2

Random uncertainty in a population

As mentioned previously if you take a series of measurements of the same quantity you have what is referred to as a populations of measured values. These will have an uncertainty that is due to random variations in the measurement. The example was given of measuring the final velocity of a ½ kg mass dropped 100 times. You would have a population of 100 data points with a small random variation. You will find that these measurements will distribute themselves in a 'bell' curve, also known as a normal or Gaussian distribution. We can use some basic statistics to find the nominal value and the statistical uncertainty of that value.

The mean (also known as the average) value of the population is used as the nominal value and can be found by using the following:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Where X_i is a individual measured value, N is the number of measurements in the population and $\overline{X}$ is the mean.

As an example here is a sample population of measurements in cm:

2.3, 2.5, 3.1, 1.9, 2.2, 2.7 and 2.6.

The mean of our sample population is calculated:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{2.3 + 2.5 + 3.1 + 1.9 + 2.2 + 2.7 + 2.6}{7} = 2.471\text{cm}$$

We can now estimate the uncertainty by using the standard deviation σ_x

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i-1}^{N}\left(X_i - \overline{X}\right)^2}$$

For our sample population σ_x = 0.386 cm.

Putting these together our nominal value and uncertainty are 2.5 ± 0.4 cm.

We know that for such a distribution, 68% of the measurements are predicted to lie in the range $\overline{X} + \sigma_x$ to $\overline{X} - \sigma_x$ (within 1 σ). So 68% of the measurements in this population are between 2.1 cm and 2.9 cm. If we take another measurement we have a 68% chance that it will fall in this range. A 1.5 σ range gives us 87%, a 2 σ range gives us 95%, and a 3 σ range gives us 99.7% of the population.

Comparisons

Now that you have a nominal value and an uncertainty the final step in numerical analysis is to determine the level of agreement between your experimental result and the theoretical or expected value. This is done in a variety of ways depending on the type of data that you have, but all entail comparison with

How many sigma?

The number of sigma (σ) the experimental mean is away from a known theoretical or expected value is a very useful means of determining your experimental accuracy.

Number of Sigmas	Quality of Agreement
< 1 σ	Very Good
1 σ	Good
2 σ - 3 σ	Fair
> 3 σ	Poor

Discrepancy

Discrepancy (also know as the absolute difference) is the positive difference between the experimental result that you achieve and the theoretical value. The following formula will allow you to calculate discrepancy:

Discrepancy = |Experimental Value — Theoretical Value|

The experimental value is the nominal value you determined by one of the previous methods. Very often the worst case discrepancy is used. This is found by taking either the upper or lower bound of the experimental range (which ever is further away from the expected or theoretical value) and using that as the experimental value is the above formula.

Comparison with uncertainty

This is done by comparison of your uncertainty to the discrepancy.
If the discrepancy is:

≤ uncertainty then you are in <u>Agreement</u>.
> uncertainty but ≤ 3 times uncertainty, you are in <u>Marginal Agreement</u>.
> 3 times uncertainty, then you are in <u>Disagreement</u>.

The idea here is to show if the discrepancy can be accounted for by the uncertainty present in the measurement.

Fractional error

This comparison is of the discrepancy to the theoretical or known value by finding the ratio of the two.

Fractional Error = |Experimental — Theoretical| / Theoretical

Percent error

The fractional error is often shown as a percent.

% Error = Fractional Error x 100%

or

% Error = (|Experimental — Theoretical| / Theoretical) x 100%

Percent difference

When you are comparing two experimental values and there is no known value you show the percent difference between the two measurements.

$$\% \text{ Difference} = \frac{(|\text{Experimental 1} - \text{Experimental 2}|}{\frac{1}{2}(\text{Experimental 1} + \text{Experimental 2})} \times 100\%$$