

University of Massachusetts Boston - Chemistry Department
Physical Chemistry Laboratory
Introduction to Maximum Probable Error

Statistical methods describe **random** or **indeterminate** errors in experimental data and in quantities calculated from those data. Statistical uncertainties such as the standard deviation describe the distribution of results obtained by repeated measurements (*precision*). Such values are useful descriptions of the uncertainty in our results as long as we have made an accurate measurement. But how can we determine the *accuracy* of our measurements?

No number of repeated measurements will provide an estimate of the accuracy of a result. Even measurements “relatively free” from random errors are subject to **systematic** or **determinate** errors. Suppose that we measure the length of some object with a ruler. We might repeat the measurement thousands of times and obtain a highly precise and reproducible average length of 7.50 ± 0.07 inches. Suppose that the ruler (a less than perfect physical object) was manufactured to a tolerance of $\pm 2.0\%$. What we read as 7.50 inches might be as much as 7.65 inches or as little as 7.35 inches or anywhere in between. The $\pm 2.0\%$ uncertainty represents a systematic error in the measurement. No statistical analysis can account for this error.

Systematic errors result in a discrepancy between a result derived from measured quantities and the “true”, “actual”, “authentic”, “correct” result. Of course, it is impossible to measure the “true” result. (Perhaps the very concept of a “true” experimental result is a meaningless abstraction.) In practical terms the evaluation of systematic errors allows comparison of different results from different experimental sources. More importantly, error analysis calculations are an essential part of experimental design.

It is important that the experimenter is very careful in designing the experiment to minimize systematic errors. We will learn how to measure the effects that systematic have on the accuracy of our results. This process is not difficult in simple cases, but as experiments become more complex, analysis of systematic error becomes very complicated and “fuzzy”. In general, in a well designed experiment random errors are large compared to systematic errors. If this is not the case, often improvements in the experimental design can be made in efforts to further minimize systematic errors. Therefore, a detailed analysis of systematic error is an important part of experimental design.

Analysis of systematic error

Suppose that some quantity Q is to be determined by experiment. In general, the value of Q is not measured directly in a simple experiment. Instead, one or several quantities which we denote as x_1, x_2, \dots, x_n (or collectively as x_i) are measured and a value of Q is calculated from these by means of some equation or by a graphical procedure. It is usually possible to estimate the uncertainties of direct x_i measurements from information such as instrument specifications, calibration experiments and glassware tolerances. Our objective here is to show how to relate the estimated errors Δx_i in the measurements x_i to the error in Q . This calculational procedure is called error propagation. This procedure estimates the so-called *maximum probable error* in Q which we shall denote by ΔQ . We call this a maximum probable error because it assumes that contributing errors from separate measurements all combine to perturb the result in the same direction; we do not assume that an error due to any one measurement cancels

any error from other measurements. (This is a simple application of Murphy's Law - if anything can possibly go wrong, it will.) With this assumption and if all contributing errors are properly estimated and combined, the "true" value of Q should fall in the range $Q - \Delta Q$ to $Q + \Delta Q$.

This is so important that I'll say it again. We have determined some quantity, Q. Is this value "correct"? No. However, I can be sure that the correct value of Q, whatever it is, lies in the range $Q - \Delta Q$ to $Q + \Delta Q$.

The error propagation calculation is fairly simple. Suppose that we have measured values for x_1, x_2, \dots, x_n . We use these data to calculate some quantity Q. We use the symbol Q to indicate our best value for the quantity. We recognize that each x_i value may be in error and assume that we know the maximum bounds of the x_i errors. For example, x_1 might represent the volume delivered by a 10 mL pipet. Perhaps the pipet used in the experiment was manufactured with a tolerance, Δx_1 , of ± 0.02 mL. If the pipet was used correctly, the delivered volume certainly was within the range 9.98 to 10.02 mL. (Well, almost certainly; nothing is really certain.) We now consider the perturbation (error) in Q caused by the uncertainty in x_1 . We suppose that the "true" value of x_1 might be as large as $x_1 + \Delta x_1$ (10.02 mL) and recalculate the value of Q using the new x_1 value, 10.02 mL. From the calculation we obtain a new result which we denote by Q_1 whose value differs from Q by an amount $\Delta Q_1 = Q_1 - Q$. The value ΔQ_1 represents the uncertainty generated in Q as a consequence of the uncertainty in x_1 . Next we consider the perturbation in the result caused by uncertainty in x_2 . We recalculate the result taking x_2 as its maximum value $x_2 + \Delta x_2$. In this calculation the values of $x_1 \dots x_n$ remain at their original values. We find a new result Q_2 which differs from Q by an amount, $\Delta Q_2 = Q_2 - Q$. The recalculation procedure is repeated for each x_1, x_2, \dots, x_n to account for all possible sources of error in Q. Finally, the maximum probable error in Q is given by:

$\Delta Q = |\Delta Q_1| + |\Delta Q_2| + \dots + |\Delta Q_n| = \sum |\Delta Q_i|$, where each term represents the absolute value of $\Delta Q_1, \Delta Q_2$ etc. It is important to note that each ΔQ calculation should be made to one or two extra significant figures than is customary. We illustrate the procedure below.

EXAMPLE 1

We wish to weigh out some NaCl using a balance which supposedly is accurate to ± 0.10 g. First we weigh an empty beaker and record a 15.0 g weight. Next we add NaCl to the beaker and record a weight of 20.0 g. What weight of NaCl is in the beaker and what is its maximum error?

SOLUTION

The weight of NaCl is the difference between the weight of the (beaker + NaCl) and the weight of the empty beaker, 5.0 g of NaCl. However, both the (beaker + NaCl) and empty beaker weights might be in error by as much as 0.10 g. We will call the (beaker + NaCl) = F and the empty beaker E. We calculate the weight of NaCl, $Q = F - E$, from $Q = F - E$ and suppose first that F might be too high by 0.10 g. Then $Q_1 = 20.10 - 15.00 = 5.10$ g; $\Delta Q_1 = 5.10 - 5.00 = 0.10$ g. Next, we find the error in Q due to the possible 0.10 g error in E. $Q_2 = 20.00 - 15.10 = 4.90$ g; $\Delta Q_2 = 0.10$ g. $\Delta Q = |\Delta Q_1| + |\Delta Q_2| = 0.20$ g. In short, we have weighed out 5.0 ± 0.2 g NaCl and may be assured that the actual weight of NaCl is between 4.8 and 5.2 g.

EXAMPLE 2

The 5.0 g sample of NaCl above is dissolved and diluted to 40.0 mL in a 50 mL graduated cylinder. Calculate the concentration of the solution in g/mL units and its maximum probable error. The weight is believed accurate to ± 0.2 g and the cylinder is accurate to $\pm 2.0\%$ of capacity or ± 1.0 mL.

SOLUTION

The concentration of NaCl solution (Q) is $5.0/40.0 = 0.125$ g/mL. There are two possible error sources: the weighing and the volume. We evaluate the weighing error by using the extreme value of the NaCl weight $5.0 + 0.2 = 5.2$ g to recalculate the concentration.

$$Q_1 = 5.2/40.0 \text{ g} = 0.130 \text{ g/mL}$$

$$\Delta Q_1 = Q_1 - Q = 0.130 - 0.125 = 0.005 \text{ g/mL}$$

We evaluate the volume error by using its extreme value to recalculate the concentration.

$$Q_2 = 5.0/41.0 = 0.122 \text{ g/mL}$$

$$\Delta Q_2 = Q_2 - Q = 0.122 - 0.125 = -0.003 \text{ g/mL}$$

The uncertainty in NaCl concentration is:

$$\Delta Q = |\Delta Q_1| + |\Delta Q_2| = 0.005 + 0.003 = 0.008 \text{ g/mL.}$$

Thus the concentration of NaCl is $0.125 (\pm 0.008)$ g/mL. This value means that if the balance worked properly and if the cylinder was made properly and we used them properly, we can be pretty sure that the actual NaCl concentration is in the range 0.125 ± 0.008 g/mL, that is between 0.117 and 0.133 g/mL.

Comment:

For the above example it is possible to eliminate the systematic errors that are limiting the accuracy of the result. For instance, one could calibrate the scale with very accurate weights (± 0.0001 g). One could calibrate the volume of the graduated cylinder by precisely weighing the graduated cylinder filled to the mark with water. If one performs this measurement several times one might find that it is really a 50.81 ± 0.08 ml graduated cylinder. Thus, the graduated cylinder is now accurate to 0.16 %. It is often possible to reduce the systematic errors to the point where they become insignificant compared to the random errors inherent in the experiment. Then, we can treat our data statistically with a clear conscience.

EXAMPLE 3

A pH meter, believed accurate to ± 0.02 pH units, is used to measure a 0.100 ± 0.005 M solution of acetic acid in water. The pH reading is 2.87. Neglect activity effects and calculate K_a for acetic acid and its uncertainty.

SOLUTION

We will use the equation

$$K_a = \frac{[H^+]^2}{C - [H^+]}$$

and take $[H^+] = 10^{-\text{pH}}$ and $C = 0.100$ M.

$$K_a = \frac{(10^{-2.87})^2}{0.100 - (10^{-2.87})} = 1.8_5 \times 10^{-5}$$

We consider the uncertainty in K_a due to the pH error of ± 0.02 and adjust the pH value to 2.89.

$$K_a = \frac{(10^{-2.89})^2}{0.100 - (10^{-2.89})} = 1.6_8 \times 10^{-5}$$

$$\Delta(K_a)_1 = 1.6_8 \times 10^{-5} - 1.8_5 \times 10^{-5} = 0.1_7 \times 10^{-5}$$

The uncertainty in K_a due to the error in C is found by adjusting the value of C.

$$K_a = \frac{(10^{-2.87})^2}{0.105 - (10^{-2.87})} = 1.7_8 \times 10^{-5}$$

$$\Delta(K_a)_2 = 1.7_6 \times 10^{-5} - 1.8_5 \times 10^{-5} = -0.0_9 \times 10^{-5}$$

The uncertainty in K_a is $0.1_7 \times 10^{-5} + 0.0_9 \times 10^{-5} = 0.2_6 \times 10^{-5}$. K_a is reported as $1.85 (\pm 0.26) \times 10^{-5}$.

The method outlined above will always provide correct estimates of the maximum probable error no matter how complex the calculation for Q may be. The recalculation procedure represents a slow but safe method of error propagation. It relies on only one fundamental assumption: that each x_i represents an independent measurement. This is obviously the case in the first examples where the measurement of NaCl mass was made with two separate weighings using a balance and the volume measurement employed a graduated cylinder. Likewise, in the third example the pH meter uncertainty represents an error in an electronic pH meter while the concentration error derives from some other experiment which might rely on weighing or a titration.

In some cases the recalculation method is quite tedious and it is convenient to recognize some special cases which are frequently encountered in error propagation calculations.

SHORTCUTS

1. Multiplication or Division by an Exact Quantity

Suppose that an experimental number, x , is multiplied (or divided) by some exact quantity. The uncertainty in the result is the uncertainty in x multiplied (or divided) by the same exact quantity.

Example: The radius of a circle is measured as 1.00 cm with a maximum uncertainty of 0.01_0 cm. The circumference is $2\pi(1.00 \pm 0.01_0)$ cm = 6.28 ± 0.06 cm.

2. Addition and Subtraction of Experimental Numbers

Suppose some result is calculated as the sum or difference between two numbers. The

uncertainty in the sum (or difference) is the sum of uncertainties in the individual values.

Examples (similar to Example 1)

a. A 10.0 ± 0.1 mL portion of water is added to a bottle. A second portion of 15.0 ± 0.2 mL of water is added to the same bottle. The amount of water in the bottle is 25.0 ± 0.3 mL.

b. An empty beaker weighs 15.0 ± 0.1 g. Some solid is added and the beaker reweighed. The combined weight of solid and beaker is 20.0 ± 0.1 g. The beaker contains 5.0 ± 0.2 g. of solid.

3. Multiplication and Division of Experimental Numbers

Suppose that some result is calculated as the product or ratio of two numbers. The **relative** error in the product (or ratio) is the sum of the **relative** errors of the numbers.

Examples (similar to Example 2)

a. A 10.0 ± 0.2 mL portion of a liquid weighs 20.0 ± 0.1 g. The density of the liquid is $2.00 \text{ g}/10.0 \text{ mL} = 2.00 \text{ g/mL}$. The relative uncertainty in the mass is $0.1 \text{ g}/20.0 \text{ g} = 0.005$ or 0.5% . The relative uncertainty in the volume is $0.2 \text{ mL}/10.0 \text{ mL} = 0.02_0$ or $2_0\%$. The relative uncertainty in the density is $2_5\%$ or $0.02_5 \times 2.00 \text{ g/mL} = 0.05 \text{ g/mL}$. The density is $2.00(\pm 0.05)\text{g/mL}$.

b. A sample is analyzed for chloride by titration with 32.50 ± 0.10 mL of $0.1000 \pm 0.0002 \text{ M AgNO}_3$. What weight of chloride is present in the sample?

The weight of chloride present is $(32.50 \text{ mL} \times 0.1000 \text{ M}) \times 35.453 = 115.2 \text{ mg}$. The relative error in the volume is $0.10/32.5 = 0.31\%$. The relative error in the concentration of AgNO_3 is $0.0002/0.1000 = 0.2_0\%$. The uncertainty in the weight of chloride is $0.31\% + 0.2_0\% = 0.5_1\%$; $0.5_1\%$ of $115.2 \text{ mg} = 0.6 \text{ mg}$. The weight of chloride present is $115.2 \pm 0.6 \text{ mg}$.

It is important to note that the special shortcuts may be applied only when x_i values are independent. To repeat, the x_i must come from separate measurements; they must not rely on the same measurement quantities. Consider the following example.

EXAMPLE 4

Consider Tom, Dick and Harry.

Tom is 68.0 ± 1.0 inches tall. Dick is 3.0 ± 0.5 inches taller than Tom. Harry is 5.0 ± 0.5 inches taller than Tom. How much taller is Harry than Dick?

Harry is $(5.0(\pm 0.5) - 3.0(\pm 0.5)) = 2.0 \pm 1.0$ inches taller than Dick.

Suppose we had tried to find the difference another way. Since Tom is $T = 68.0 \pm 1.0$ inches tall, Dick is $T + d = 68.0(\pm 1.0) + 3.0(\pm 0.5) = 71.0 \pm 1.5$ inches tall and Harry is $T + h = 68.0(\pm 1.0) + 5.0(\pm 0.5) = 73.0 \pm 1.5$ inches tall. The height difference between Harry and Dick is $73.0(\pm 1.5) - 71.0(\pm 1.5) = 2.0 \pm 3.0$ inches. That is, Harry is between 5.0 and -1.0 inches taller than Dick. Harry might be an inch shorter than Dick! Nonsense. The difficulty is that we have included Tom's height and its uncertainty in the calculation.

In terms of equations, we seek $h - d$ (Harry - Dick) which is $5.0(\pm 0.5) - 3.0(\pm 0.5) = 2.0 \pm 1.0$ inches. By calculating the actual heights of Dick and Harry we find Harry's height, $T + h$, and Dick's height, $T + d$. Subtracting to find $h - d$ we

find:

$$h - d = (T + h) - (T + d)$$

This equation is algebraically correct but the error analysis is incorrect because we have accounted for Tom's height and its uncertainty in two places (the first and second terms of the equation) where, in fact, Tom's height is irrelevant to the result. We have gone wrong by considering Dick and Harry's calculated heights as independent x_i . Clearly, both values, 71.0 ± 1.5 inches and 73.0 ± 1.5 , depend on Tom's height. Harry's and Dick's heights are calculated results depending on the same data and are not independent.

EXAMPLE 5

A solution is prepared by adding together 10.0 ± 0.1 moles of water and 2.00 ± 0.05 moles of ethanol. What is the mole fraction of water in the mixture and its maximum error?

Solution:

The mole fraction of water, $X_w = n_w / (n_w + n_{\text{ethanol}})$. Substituting the values given, $X_w = 10.0 / (10.0 + 2.00) = 0.833_3$. This is our best estimate. We may estimate the maximum error in this value in two ways.

THE WRONG ERROR CALCULATION:

$$X_w = (10.0 \pm 0.1) / [(10.0 \pm 0.1) + (2.00 \pm 0.05)]$$

The denominator has a value of $12.00 \pm 0.1_5$. The numerator is 10.0 ± 0.1 . The relative errors in the numerator and denominator are 1.0% and 1.3%, respectively. The relative error of X_w is 2.3% so that the error in X_w is $0.833_3 \times 2.3\% = 0.01_9$. This answer is wrong.

THE RIGHT ERROR CALCULATION:

There are two independent sources of error: n_w and n_{ethanol} . We first change the value of n_w to its extreme of 10.1 and calculate

$$(X_w)_1 = 10.1 / (10.1 + 2.00) = 0.834_7$$

This differs from the best value by $|0.834_7 - 0.833_3| = 0.001_4$.

Next we change the value of n_{ethanol} to its extreme of 2.05 and calculate

$$(X_w)_2 = 10.0 / (10.0 + 2.05) = 0.829_9$$

This result differs from the best value by $|0.829_9 - 0.833_3| = 0.000_4$. The maximum error in X_w is $0.001_4 + 0.000_4 = 0.001_8$. The mole fraction of water in the mixture is $0.833_3 \pm 0.001_8$. This is the right answer.

Why are the answers different? (Note that the errors differ by a factor of about 10) Ask Tom, Dick or Harry. The shortcut (wrong) method makes the hidden assumption that the numerator (n_w) is independent of the denominator ($n_w + n_{\text{ethanol}}$). Clearly this is not the case. The wrong assumption leads to a wrong answer.

MAXIMUM ERROR ESTIMATES OF STRAIGHT-LINE PARAMETERS

Estimating the maximum error of straight-line parameters is tricky business. Remember, by estimating the maximum error we are trying to account for any unavoidable systematic errors that might be in any of the measurements that we have made. It has nothing to do with the statistical spread in our data. In most instances, the experimenter collects all data points that make up the plot using the identical

experimental technique. If this is the case, it is likely that all of the data points will carry **the same systematic error!!!** It is shown in the example below that **under these constraints**, the intercept depends on the magnitude of the systematic error, but **the slope is independent of systematic error.**

EXAMPLE 6

Consider the x-y data listed below.

x	y
1	2.1
2	4.3
3	6
4	8.6
5	10.2

Using the function key in Excel, we find that the slope = 2.05, and the intercept = 0.09. I am assuming that you all have used most simple Excel functions before, but please feel free to ask questions about these procedures if you don't know how or don't remember.

Based on our experimental procedure, we can estimate the maximum probable error in the x data and in the y data. For instance, we may reason that the x data could have a maximum systematic error of ± 0.25 , and that the y data could have a maximum systematic error in the measurement of ± 0.5 . These are estimations of systematic errors in the measurement of x and y, respectively.

If we construct a plot with error bars, we can draw boxes around each of the points that represents the maximum errors acceptable for each of the points. You should be able to draw a line that intersects all of boxes. If a point fails to meet this criterion you can discard it!!

Back to our example. If the first x, y data point is off by +0.2 in x and -0.3 in y, then all of the other data points are also off by the same systematic error. In other words all of the data points are fraught with the same systematic error, otherwise the error would not be systematic, it would be statistical. So what happens to the slope and the intercept if we plot the extremes of the possible systematic error in x and y?

CAUTION: It is possible that the magnitude of a systematic error can sometimes be proportional to the magnitude of the measurement. In other words the systematic error does not remain constant across the range of x or y values. In these cases estimating the maximum error becomes increasingly challenging, but often still manageable.

xmax	ymin	xmin	ymax
1.25	1.6	0.75	2.6
2.25	3.8	1.75	4.8
3.25	5.5	2.75	6.5
4.25	8.1	3.75	9.1
5.25	9.7	4.75	10.7
	slope	int	
minmin	2.05	0.1025	
		-	
maxmin	2.05	0.9225	
minmax	2.05	1.1025	
maxmax	2.05	0.0775	

Careful inspection of the data would reveal that the (max,min) data gives the lowest possible intercept and the (min, max) data gives the greatest possible intercept. The systematic error in the intercept ($s_{mpe,int}$) is given by

$$[(\text{intercept})_{\text{max}} - (\text{intercept})_{\text{min}}]/2 \text{ or in this case, } (1.1025 - -0.9225)/2 = 1.01$$

$$\text{intercept} = 0.090 \pm 1.01$$

However, the slope is **independent** of systematic error. So, how do we represent the error in the slope?

Statistical error in the slope

In the above example the error in the slope is a function of the purely statistical spread in the data. One can use Excel to calculate the statistical error in the slope and intercept. Click on “tools”. Is there a “Data Analysis” option? If so, click on it. If not you must add it by clicking on “add-ins” and selecting the “Analysis Tool Pack”. Add this tool pack, then go back into “tools” and the “Data Analysis Option” should now be listed. Once you click on it, scroll down to “regression”, click on it, and hit “OK”. Input the y and x data, select the “output range” box and choose a cell to place the data that is far removed from your data. Hit OK! On the data sheet scroll down to the cell where you have chosen to place your results. Under “regression statistics” there are two pieces of important information, adjusted R^2 and standard error. The adjusted R^2 is a measure of the linear correlation of your data. The closer to 1 (or negative 1, if the slope is negative), the better.

In the second section below this you will see “intercept” and “x variable” listed (x variable is really the slope). The cells immediately adjacent to these are the actual values for the intercept and slope, and next to these are the statistical values for the errors (standard deviations) in the intercept and slope. To the right of these cells are maximum and minimum values for the slope and intercept at the 95 % confidence limit. By subtracting the minimum from the maximum and dividing by 2, we get a reasonable estimate of the error in the slope or intercept at the 95 % CL. We will generally use these values when reporting the statistical error in the slope and intercept.

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.99778253
R Square	0.99556998
Adjusted R Square	0.99409331
Standard Error	0.24966644
Observations	5

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0.09	0.261852	0.343705	0.753747	-0.74333	0.923331
X Variable 1	2.05	0.078951	25.96532	0.000125	1.798741	2.301259

$$\text{Statistical error in Intercept} = s_{\text{stat,int}} = (0.923+0.743)/2 = 0.833$$

$$\text{Intercept} = 0.09 \pm 0.83_3$$

$$\text{Statistical error in Slope} = s_{\text{stat,slope}} = (2.301-1.799)/2 = 0.251$$

$$\text{Slope} = 2.05 \pm 0.25_1$$

If you are unfamiliar with confidence limits see chapter 4 of Harris

Propagation of error using parameters from the least squares analysis

As another example, let's consider the typical spectroscopic experiment. You measure the absorbance of standards to create a Beer's law plot, which is then used to calculate the concentration of an unknown.

$$C = (\text{Abs} - \text{intercept})/\text{slope}$$

Propagation of maximum probable error.

As stated above only the intercept is affected by systematic error. Therefore, to calculate the error in the unknown concentration resulting from possible systematic errors, we need only plug in the extreme values for the intercept into the above equation to find the extreme values of C. Dividing the range by 2 gives us the maximum probable error in C.

$$\text{Max. prob. Error} = s_{\text{mpe},C} = [C(\text{max}) - C(\text{min})]/2$$

Propagation of statistical error

We will use the maximum and minimum values of the slope and intercept at the 95 % confidence limit to find the error at any given x value, ie. C.

How do we calculate the statistical error in the unknown concentration? This is more complicated than it looks at first glance. On instinct, one might first use the law of relative errors. Recall (or learn here) that, **while systematic errors are simply additive, statistical errors add in quadrature.** In this case, that would mean:

$$(s_{\text{stat},C}/C)^2 = (s_{\text{stat,int}}/\text{int})^2 + (s_{\text{stat,slope}}/\text{slope})^2$$

However, the statistical errors in the slope and intercept are not independent of each other (the Tom, Dick, and Harry example all over again)!

In actuality, the intercept must be at a minimum if the slope is at a maximum. The corollary to this is also true: the intercept must be at a maximum if the slope is at a minimum, so long as your data points consist of positive numbers.

To approximate the statistical error, we will take advantage of Excel's statistical output. We use the below equation,

$$C = (\text{Abs} - \text{intercept})/\text{slope}$$

plugging in the lower confidence limit of the slope and the higher confidence limit of the intercept to calculate C(extreme1), and plugging in the higher confidence limit of the slope and the lower confidence limit of the intercept to calculate C(extreme2).

Statistical error in C is then given by,

$$\text{Statistical error} = s_{\text{stat},C} = (C(\text{extreme1}) - C(\text{extreme2}))/2$$

Combining the maximum probable error and the statistical error to get an overall error.

As a good estimation of the overall error, these two errors are simply additive.

$$S_{\text{overall}} = s_{\text{mpe}} + s_{\text{stat}}$$

Using s_{mpe} and s_{stat} to guide experimental design.

More importantly, perhaps, is the comparison of s_{mpe} and s_{stat} . If $s_{\text{mpe}} \gg s_{\text{stat}}$, it is likely that you can design the experiment in a better way that will reduce s_{mpe} . However, if the statistical spread in your data is the greatest source of error, then you are likely to have designed the experiment well.

A FINAL COMMENT

The most important application of error propagation calculations is not in reporting the accuracy of some quantity determined by experiment. Instead, these calculations are critical to the design of the experiments themselves. Consider the experiment described in Example 2. The uncertainty in the concentration of a NaCl solution depended on uncertainties in both the NaCl weight and solution volume. Most of the possible error in the concentration resulted from the weighing error of ± 0.2 g. (We found $\Delta Q_1 = 0.005$ g/mL compared to $\Delta Q = 0.008$ g/mL.) Clearly, by using a better balance, accurate to perhaps 0.01 g or 0.02 g, the accuracy of the concentration value would be substantially improved. However, even if a really good balance, accurate to 0.00001 g were used, the 0.003 g/mL error due to volume uncertainty would remain. In other words it would be pointless to use a fancy balance if we continue to use a graduated cylinder to measure the volume.

Next, consider the experiment of Example 3. A K_a value is determined by measuring the pH of a dilute acetic acid solution. Both the pH uncertainty and uncertainty in the acid concentration contributed to the error in K_a . About 2/3 of the uncertainty derived from the ± 0.02 pH error and about 1/3 from the $\pm 5\%$ error in concentration. Certainly we could slightly improve the K_a estimate by making an acid solution with more accurately known concentration, say $\pm 1\%$. However, there would be no point in working hard to make a $\pm 0.2\%$ acetic acid solution. The main error source in the experiment is the pH measurement and an improved result will rely on reducing the ± 0.02 pH unit uncertainty. The error propagation calculation is essential in identifying the sources of uncertainty **BEFORE THE EXPERIMENT IS PERFORMED**. A little thought and calculation can prevent months of useless work and misleading results.