CHAPTER 1. INTRODUCTORY MATERIAL

Analytical chemistry is that branch of chemistry that is concerned with determining the composition of materials. This involves both identifying what compounds are present in a sample (qualitative analysis) and specifying the quantity of each compound in the sample (quantitative analysis). It is traditional that introductory analytical texts emphasize quantitative analysis. This text, which will follow tradition, presents an introduction to the theory and practice of quantitative chemical analysis, both classical and simple instrumental analysis. Some introductory material on statistics and stoichiometry will be presented first. However, before beginning with this it should be noted that quadratic equations will be encountered quite often in this material and that the use of calculators is highly recommended for problem solving. There is a short section on each of these topics in Appendix A, which you should read and understand before attempting the problems included with each chapter.

1.1 STATISTICAL TREATMENT OF DATA

The final result of a quantitative chemical analysis is almost always one or more numbers which indicate the concentrations of the compounds in the sample. There is always some uncertainty associated with each operation or measurement in an analysis and thus there is always some uncertainty in the final result. Knowing the uncertainty is as important as knowing the final result. Having data that are so uncertain as to be useless is no better than having no data at all. Thus there is a need to describe and reduce, if necessary and possible, this uncertainty. Statistics is of great help in this regard. To start, two terms, which are often used incorrectly, are defined.

Accuracy - the closeness of the experimental result to the actual value (the correct answer).

Precision - the reproducibility of a data set; a measure of the ability to obtain the same number (not necessarily the correct number) in every trial.

Precision and accuracy are *not* necessarily related, as described below. However, it is normal to place more "faith" in the results of an analyst who has good precision, since good precision usually means greater attention to detail.

1.1.1 TYPES OF ERRORS

Good accuracy should be the object of all chemical measurements. The question thus arises, how is accuracy measured? This topic is best introduced by discussing errors, since a measurement will be accurate if there are no errors. There are three general types of errors:

1) **Blunders** - these are mistakes that are obvious to the analyst and that are unlikely to be repeated when another analysis is performed.

An example of a blunder is spilling a sample on the laboratory bench.

2) **Determinate errors** - these errors, often referred to as **"bias"**, are also mistakes, but are much less obvious to the analyst and are quite likely to be repeated when other analyses are performed. In theory many determinate errors, if recognized, could be eliminated by improving technique.

An example of a determinate error would be the use of an analytical balance that has gone out of calibration to weigh a sample. If the balance weighs low (the sample is heavier than the balance indicates) the analytical result will always be high. Thus determinate errors are unidirectional - they always produce an error in the same direction.

3) **Indeterminate errors** - these errors, often referred to as **"random errors"**, are caused by the need to make estimates in the last figure of a measurement, by noise present in instruments, *etc*. Such errors can be reduced, but never entirely eliminated.

An example of indeterminate error is the variation in the values taken from a buret due to the need to estimate the last significant figure. The observer would certainly say, referring to Figure 1-1, that the volume delivered is between 40.2 and 40.3 mL. However, one can and should do better than this. Some observers might say that the volume was 40.24, some 40.25, and others 40.26 mL. This uncertainty in the last place is an example of indeterminate error. It is better to report the volume as 40.26 mL rather than to report its being between 40.2 and 40.3 mL, even though it is realized that the 0.06 is an intelligent estimate. Indeterminate errors are bidirectional, being random in nature they are as likely to produce a high result as a low result.

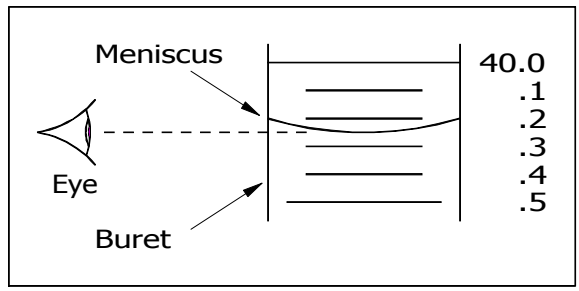

Figure 1-1

Note that determinate error may also be involved in estimating the last figure of a measurement. In the case being discussed parallax error (looking at the meniscus at an angle) and number bias (selecting some numerals, *e.g.* 0 or 5, more often than others in the estimate) are two examples of determinate error.

Since determinate errors are usually unidirectional, they will always cause an error in the final result, even if many measurements are averaged. Indeterminate errors, being bidirectional, will tend to cancel if enough measurements are averaged. These ideas are perhaps best illustrated by a figure.

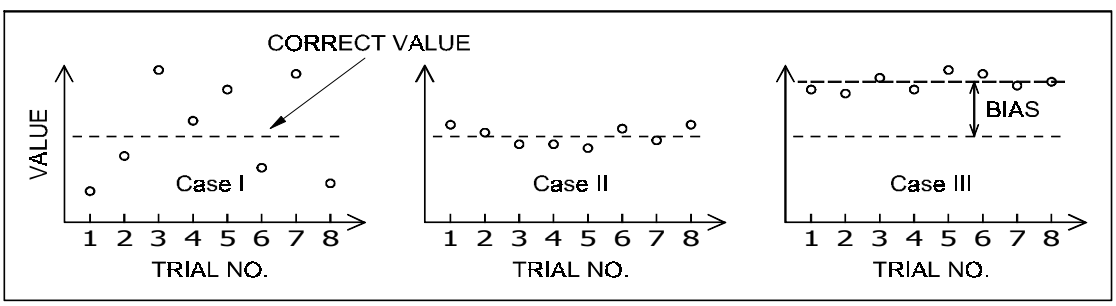

Figure 1-2

In Case I shown in Figure 1-2, the precision is poor, but the accuracy will be good if enough data are taken and averaged, since the errors are approximately equally distributed about the correct answer. In Case II the precision is good and the accuracy is also good. In Case III the precision is good, but the accuracy is poor due to a determinate error. Cases I and II illustrate the advantage of good precision. In both cases an accurate answer will be obtained if enough data are taken and averaged. However, it is not always possible to take a large number of data. If it is only possible to take a small number of data, say three, the average of such a small set of data with good precision (Case II) will give a better result than a similar average of a set with poor precision (Case I). The following summary can be given:

If there are no errors, the answer will be accurate (this is almost never the case).

If there are no determinate errors and if a *large* number of measurements are made and the average calculated, the average will be accurate.

If determinate error is present, there is usually no way to obtain the correct answer (statistics is no help).

With regard to the second statement above, how large is large? The answer is usually thirty or more. This number of measurements is almost never made in practice. However, it is still much better to make more than one measurement and take the average, even if it is only *an estimate* of the correct answer, than it is to rely on a single value. In most analytical procedures it is customary to make two or three separate measurements, except when first establishing a procedure, in which case more measurements are made.

With regard to the third statement above, since statistics is of no help when determinate errors are made, how does one know if such errors are present? The usual procedure is to analyze a standard (a sample of accurately known composition) and compare the analytical results with the known sample composition. If the values are the same, bias is assumed to be absent. Many organizations, *e.g*. the National Research Council of Canada and the National Institute of Standards and Technology in the U.S.A., provide such standards. Another procedure is to analyze the sample using two different and independent methods of analysis. If the two methods produce the same results, bias is again assumed to be absent. Unfortunately, these and other methods all have flaws and it is impossible to be *absolutely* certain whether determinate error really exists. In other words it is impossible to know with absolute certainty the correct answer.

1.1.2 STATISTICAL PARAMETERS.

In statistics we usually distinguish between parameters that pertain to a population and parameters that are obtained from a finite data set. A population would result if a very large number of measurements (strictly an infinite number) were made. A finite data set results when just a few measurements are made and is what we usually work with. A finite data set is a subset of the population. The parameters calculated from a finite data set, usually represented by "English" characters, are only estimates of the population parameters, usually represented by Greek characters. It is the population parameters that one would like to obtain, if possible. Much of statistics is involved with trying to make inferences about the population parameters from the data set parameters.

Since there is always uncertainty involved in all measurements, as discussed above, a data population or a data set will consist of a distribution of values rather than a single value. Such distributions have characteristics that are described by statistical parameters. The first such characteristic to be investigated is the "central tendency" of a distribution. Many distributions have some values that are more often observed than other values and these are usually toward the centre of the distribution. Three methods of defining the "central tendency" are listed below.

The mean

The mean or average is calculated by summing all the values, x_i, in the set and dividing by the number of trials, N.

$$\text{Mean} = m = \bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N} \qquad (1\text{-}1a)$$

or

$$m = \frac{\sum_{i=1}^{i=N} x_i}{N} \tag{1-1b}$$

As $N \to \infty$, $m \to \mu$, where μ is the population mean. In the *absence of determinate error*, μ is the correct answer, and as N becomes larger the estimate, m, of the correct answer gets better.

The median

The median is the "central point" in a data set. Half of all the values in a set will lie above the median, half will lie below the median. If the set contains an odd number of data points, the median will be the central point of that set. If the set contains an even number of points, the median will be the average of the two central points.

The mode

The mode is the value that occurs most often (the value with the highest probability). The mode is quite often not obvious in a finite data set where no two values are the same.

In populations where errors are evenly distributed about the mean (the distribution is symmetrical), the mean, the median, and the mode will all have the same value. In this text, the mean will be the most useful parameter for indicating the "central tendency" of a data set or population.

Accuracy and how one attempts to achieve and measure it have been discussed. The topic of precision comes next. How does one measure precision? Obviously one must have made more than one measurement to get an estimate of precision. The characteristic of a distribution that relates to precision is the variability. In data distributions that arise from physical measurements the variability is a direct measure of the precision of the measurement. The parameters used to describe the variability are listed below.

The standard deviation

The standard deviation (and those parameters related to it) is the only statistically acceptable measure of precision and is the only measure to be discussed in detail here. It is calculated from the mean and the values of the trails in the data set according to:

$$s = \sqrt{\frac{(x_1 - m)^2 + (x_2 - m)^2 + \cdots + (x_N - m)^2}{\text{D.F.}}} \tag{1-2}$$

The values of $(x_i - m)$ are the deviations from the mean and D.F. stands for the degrees of freedom. The degrees of freedom are usually one less than the number of data in the set, $(N-1)$. The reason for this is that one degree of freedom is lost when calculating the mean of the data set. In a data set of ten, for example, the tenth datum can be calculated from the other nine and the mean, indicating that there are only nine independent data left after the mean has been calculated. Thus

$$s = \sqrt{\frac{\sum (x_i - m)^2}{N - 1}} \tag{1-3}$$

As with the mean, as $N \to \infty$, $s \to \sigma$, where σ is the population standard deviation. s is only an estimate of σ, but as N gets larger, that estimate gets better.

The relative standard deviation

Often the relative standard deviation is more useful in a practical sense than the absolute standard deviation (given above), since it immediately provides an idea of the precision of the data set relative to its individual values. There are several ways of expressing the relative standard deviation:

$$\text{Rel. std. dev.} = s_r = \frac{s}{m} \tag{1-4}$$

$$\text{Rel. std. dev in \%} = s_r \times 100 = \frac{s \times 100}{m} \tag{1-5}$$

$$\text{Rel. std. dev. in ppt} = s_r \times 1000 = \frac{s \times 1000}{m} \tag{1-6}$$

The relative standard deviation in percent is often referred to as the *coefficient of variation*. Note that the absolute standard deviation has the same units as the mean, whereas the relative standard deviation has no units. In some analytical determinations the relative standard deviation is reported in parts per thousand (ppt).

The variance.

The variance is another method of describing variability and is often preferred by statisticians because variances are directly additive (as will be seen later). The variance is simply the square of the standard deviation.

$$\text{Variance} = v = s^2 \tag{1-7}$$

The range

The range or spread is simply the difference between the largest and smallest values in a data set.

1.1.3 THE NORMAL DISTRIBUTION

Many different data distributions are found in nature. However, it is believed that a large number of data types, including many types of chemical measurements whose variations are due to random errors, follow a Gaussian distribution. This is a symmetrical, bell-shaped distribution. Since so many data types follow this distribution it is usually referred to as the normal distribution and the mathematical equation, Eqn. (1-8), that describes this distribution is called the normal error curve.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} \tag{1-8}$$

In this equation *f(x)* is the frequency distribution function or the probability density function, x represents the trial values, μ is the population mean, and σ is the population standard deviation. The frequency distribution function gives the relative frequency at which any value of x will be observed. The relative frequency is a function of the standard deviation and the position of x with respect to the mean. A graph of the normal distribution is shown in Figure 1-3. In this figure the mean is taken as zero and the standard deviation as unity. This is the standard representation of the normal error curve.

The ability to describe the normal distribution mathematically using Eqn. (1-8) is very useful. From Figure 1-3 it is obvious that the mean of a normally distributed data population is the value that

occurs most often (has the highest frequency). Since the distribution is symmetrical, the mean, the median, and the mode will all be the same. It is less obvious that inflection points occur at $\mu \pm \sigma$. The area under the curve between x_1 and x_2, as determined by integration, gives the probability that any trial value, x, will be found between x_1 and x_2. The area between $+\infty$ and $-\infty$ is 1.0000, as it should be, since the probability of finding any value of x in this range must be unity. The area between μ and $\mu+\sigma$ is 0.3413, which means that 34.13% of all x values in a normally distributed population fall between the mean and the mean plus one standard deviation. Since the distribution is symmetrical, 34.13% of all values will fall between the mean and the mean minus one standard deviation. Thus 68.26% of all values of x will fall within $\pm 1\sigma$ of μ. Likewise it can be shown that 95.44% of all values of x will fall within $\pm 2\sigma$ of μ and 99.74% will fall within $\pm 3\sigma$ of μ. Thus, in a normally distributed population it will be very unlikely (the chances are less than 3 in 1000) that any value of x will be outside the $\pm 3\sigma$ limits.

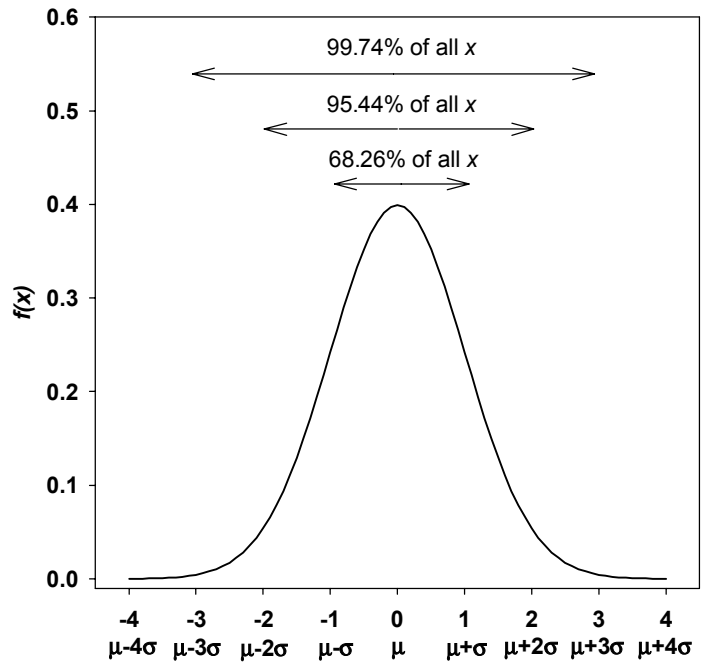

Figure 1-3

1.1.4 THE CONFIDENCE INTERVAL

One would often like to know how close the mean of a data set is likely to be to the correct answer (the population mean), assuming that there is no determinate error. Imagine not just one data set mean, but a set of means each calculated using N datum points taken at random from the population of data. This set of means would have its own mean, μ_m, and standard deviation, σ_m. Now $\mu_m = \mu$ (the population mean) and $\sigma_m < \sigma$ (the population standard deviation), and in fact it can be shown that $\sigma_m = \sigma/N^{1/2}$.

Assuming that the data populations are normally distributed (usually a good assumption), the characteristics of the standard deviation allow it to be used to estimate how close a given value is likely to be to the mean. For example, any datum in the set will be within $\pm 1\sigma$ of μ 68.3% of the time, within $\pm 2\sigma$ of μ 95.5% of the time, and within $\pm 3\sigma$ of μ 99.7% of the time (see above). If one knew the population standard deviation, σ, one could calculate the standard deviation of the set of means, σ_m. This could then be used to estimate how close the calculated mean, m, of the data set was to the mean of the means, μ_m, which is also the population mean, μ. For example one would state that

m is within $\pm k\sigma_m$ of μ with z% confidence, or

m is within $\pm k\sigma/N^{1/2}$ of μ with z% confidence

If k equals 2, then z equals 95.5. Obviously as the confidence, z, increases, k must also increase.

A problem exists, however, since one normally does not know the population standard deviation, σ. Instead one only knows s, the *estimated* population standard deviation, as calculated from the data set, Eqn(1-3). Recall that $s \rightarrow \sigma$ only as $N \rightarrow \infty$. Statisticians have developed a method to "correct" for not knowing σ. Instead of using k in the equation above, they use t, called "student's t", where $t > k$,

$t \rightarrow k$ as $N \rightarrow \infty$, and the value of t will depend both on N (actually N-1, the degrees of freedom) and the desired confidence. Values of t are given in tables, one such table being found in Appendix B. Thus the final result is:

m is within $\pm ts/N^{1/2}$ of μ with z% confidence, or

$$\text{C.I.} = m \pm \frac{ts}{\sqrt{N}} \quad \text{with } z\% \text{ confidence} \tag{1-9}$$

An example will illustrate the above points.

Assume that the following values were obtained in the analysis of the weight of iron in 2.0000 g portions of a sample: 0.3791, 0.3784, 0.3793, 0.3779, and 0.3797 g.

	x_i (g)	$(x_i-m)^2$ (g^2)				
	0.3791	$(0.3791-0.37888)^2$	=	4.84×10^{-8}		
	0.3784	$(0.3784-0.37888)^2$	=	2.30×10^{-7}		
	0.3793	$(0.3793-0.37888)^2$	=	1.76×10^{-7}		
	0.3779	$(0.3779-0.37888)^2$	=	9.60×10^{-7}		
	0.3797	$(0.3797-0.37888)^2$	=	6.72×10^{-7}		
Σx_i =	1.8944			2.09×10^{-6}	=	$\Sigma(x_i-m)^2$

The mean = m = 1.8944g/5 = 0.37888 g

The degrees of freedom = DF = (N-1) = 4

The standard deviation = $s = \{(2.09\times10^{-6}g^2)/4\}^{1/2}$ = 0.00072 g

The variance = $s^2 = (2.09\times10^{-6}g^2)/4 = 5.2\times10^{-7}\ g^2$

Rel. std. dev = s_r = 0.00072g/0.37888g = 0.0019

Rel. std. dev. in % = 0.0019(100) = 0.19%

Rel. std. dev. in ppt = 0.0019(1000) = 1.9 ppt

To easily see the range and median it is convenient to order the data in terms of increasing or decreasing values. This results in: 0.3779, 0.3784, 0.3791, 0.3793, and 0.3797 g. Since this data set has an odd number of trails, the median is simply the middle or 3rd datum, 0.3791 g. Note that for a finite data set the median and mean are not necessarily identical. The range is 0.3797-0.3779 g or 0.0018 g.

From the table of t values in Appendix B, the value of t for 4 degrees of freedom and 80% confidence is 1.533, while that for 95% confidence is 2.776. Thus $ts/N^{1/2}$ for 80% confidence is (1.533)(0.00072)/2.24 or 0.00050. The 80% confidence interval is 0.3789±0.0005 g. The 95% confidence interval can likewise be shown to be 0.3789±0.0009 g. Thus the population mean (and hopefully the correct answer) will lie between 0.3784 g and 0.3794 g 80% of the time and between 0.3780 g and 0.3798 g 95% of the time.

One could report the answer in one of several ways:

0.3789g ±0.0007g, using the absolute standard deviation
0.3789g with a rel. std. dev. of 0.0019
0.3789g ±0.19%
0.3789g ±1.9ppt

Note that the standard deviation is listed above, not the confidence interval. If one wishes to report the confidence interval, then it must be directly labeled as the confidence interval. It is normally assumed that the number following a "±" will be the standard deviation. It is tempting to report the answer as 0.37888g ±0.00072g, but this should not be done for two reasons: (1) the rules for significant figures do not allow this (it is assumed that you are familiar with these rules) and (2) the answer itself demonstrates how reporting should be done. The standard deviation indicates that there is an uncertainty of 7 in the fourth place after the decimal point. Thus it makes no sense to list any figures past this point.

1.1.5 STATISTICAL TESTS.

There are many statistical tests which are useful for chemists. These will not be mentioned with one exception. Sometimes a value in a data set appears so far from the rest of the values that one suspects that that value (called an outlier) must have been the result of some unknown large error not present in any of the other trials. It would be very useful if some test were available to determine whether the outlier could be rejected on statistical rather than on arbitrary grounds. There are several such tests, only one of which will be presented.

The Q Test

The Q Test is a very simple test for the rejection of outliers. In this test one calculates a number called Q_{exp} and compares it with values, termed Q_{crit}, from a table. If $Q_{exp}>Q_{crit}$, then the number can be rejected on statistical grounds. Q_{exp} is calculated as follows:

$$Q_{exp} = \frac{|\text{questionable value - its nearest neighbour}|}{\text{range}} \qquad (1\text{-}10)$$

An example will illustrate the use of this test.

Suppose that the following data are available: 25.27, 25.32, 25.34, and 25.61. It seems as though the largest datum may be an outlier. Q_{exp} is then calculated.

Q_{exp} = (25.61-25.34)/(25.61-25.27) = 0.79

The values of Q_{crit} are then examined in statistical tables (see Appendix B). These values depend on the number of trails in the data set, in this case 4. The values are as follows:

Q_{crit} = 0.76 at 90% confidence.

Q_{crit} = 0.85 at 96% confidence.

Q_{crit} = 0.93 at 99% confidence.

Since $Q_{exp}>Q_{crit}$ at 90% confidence, the value of 25.61 can be rejected with 90% confidence. What does this mean? It means that in rejecting the datum the experimentalist will be right an average of 9 times out of 10, or that the chances of the point actually being bad are 90%. Is this the one time out of 10 that the point is

good? This is not known! When data are rejected, there is always a risk of rejecting a good point and biasing the results in the process. Since $Q_{exp} < Q_{crit}$ at the 96% and 99% levels, the datum cannot be rejected at these levels. What this says is that if one wants to be right 96 times or more out of 100, one cannot reject the datum. It is up to you to select the level of confidence you wish to use. STATISTICAL TESTS DO NOT GIVE ABSOLUTE ANSWERS, THEY JUST TELL YOU THE ODDS.

1.2 PROPAGATION OF ERROR IN CALCULATIONS

It is an understatement to say that propagation of error calculations are unexciting. However, it is useful to know how errors propagate through a calculation or a procedure, or, in other words, how to calculate the error in the result of a calculation or procedure if you know the errors of all the numbers entering the calculation or of all the steps in the procedure. Unfortunately, determinate errors (bias) propagate through a calculation differently (since they are unidirectional) than indeterminate (random) errors. In this text only indeterminate errors will be considered. Since many determinate errors are not only unidirectional but are constant, they will not add to or change the standard deviation. For this reason it will be assumed that the standard deviation of a number arises only from indeterminate errors and that the standard deviation is a measure of the indeterminate error associated with a number.

Propagation of error calculations can provide information on which step in a procedure has the most error associated with it. This will allow you to select the proper step to improve if the errors in the procedure are too high. It can also allow you to select a step which can be done more quickly (and thus less precisely) if the error in the procedure is satisfactory and is not significantly affected by the step to be changed. Finally such calculations can indicate how low each individual step's standard deviation must be to achieve a result for the overall procedure with a desired standard deviation.

Performing a propagation of error calculation for indeterminate error basically involves applying a set of "rules" to the standard deviations of the numbers entering the calculation. In the following discussion "s" will represent the estimate of the absolute standard deviation and "s_r" that of the relative standard deviation. As indicated above the rules given below apply only to indeterminate errors.

1.2.1 ADDITION OR SUBTRACTION, $y = a \pm b \pm c$:

The absolute variance of the result, y, equals the sum of the variances of a, b, and c *etc*.

$$s_y^2 = s_a^2 + s_b^2 + s_c^2 + \cdots \tag{1-11a}$$

or

$$s_y = \sqrt{s_a^2 + s_b^2 + s_c^2 + \cdots} \tag{1-11b}$$

1.2.2 MULTIPLICATION OR DIVISION, $y = ab/c$:

The relative variance of y equals the sum of the relative variances of a, b, and c.

$$(s_y)_r^2 = (s_a)_r^2 + (s_b)_r^2 + (s_c)_r^2 \tag{1-12a}$$

or

$$(s_y)_r = \sqrt{(s_a)_r^2 + (s_b)_r^2 + (s_c)_r^2} \tag{1-12b}$$

and since

$$(s_y)_r = s_y/y \quad \text{or} \quad s_y = y(s_y)_r \tag{1-13}$$

$$s_y = y\sqrt{(s_a/a)^2 + (s_b/b)^2 + (s_c/c)^2} \tag{1-14}$$

The case of addition of or multiplication by a constant that is very accurately known, *e.g.* a molecular mass or a conversion factor, makes use of the same rules but is often somewhat simpler. In this instance it is assumed that s_k, the standard deviation of the constant, k, is essentially zero. Thus

For $y = a \pm k$, $$s_y = \{s_a^2 + s_k^2\}^{1/2} = s_a$$

For $y = ak$, $$(s_y)_r = \{(s_a)_r^2 + (s_k)_r^2\}^{1/2} = (s_a)_r$$

Note, however, that in the latter case $s_y \neq s_a$. Instead $s_y/y = s_a/a$ or $s_y = (y/a)s_a = ks_a$. (This formula is slightly different for division.)

1.2.3 RAISING TO A POWER, $y = a^x$:

In this case x is assumed to be exact, *e.g.* 2 or ½.

$$(s_y)_r = x(s_a)_r \quad \text{or} \quad s_y = yx(s_a)_r \tag{1-15}$$

1.2.4 TAKING LOGS OR ANTILOGS (BASE 10):

When the result is the logarithm to the base ten of a number, $y = (\log)_{10}a$

$$s_y = 0.43429(s_a)_r \tag{1-16}$$

and when the result is the antilog to the base ten of a number, $y = (\text{antilog})_{10}a$ or $y = 10^a$

$$(s_y)_r = s_a/0.43429 \quad \text{or} \quad s_y = ys_a/0.43429 \tag{1-17}$$

The best method for learning how to apply these "rules" is to take a few examples.

Calculate the difference between the values 4.715±0.002 and 4.693±0.003 and the absolute and relative standard deviations of the result.

It is normal procedure to calculate the result first. Thus,

$y = 4.715 - 4.693 = 0.022$

Since the only mathematical operation involved in this calculation is subtraction, the applicable equation for the absolute standard deviation is Eqn. (1-11).

$s_y = \{(0.002)^2 + (0.003)^2\}^{1/2} = 0.0036 = 0.004$

The relative standard deviation is calculated using Eqn. (1-13).

$(s_y)_r = 0.0036/0.022 = 0.16$ (or 16% relative standard deviation)

In the above example note that both numbers entering the calculation are accurately known, the number 4.715 has a relative standard deviation of 0.04% and 4.693 has an rsd of 0.06%, yet the result of the calculation has an rsd of 16%. This illustrates that differences between two similar numbers are usually not very accurate (on a relative basis), even when the numbers entering the calculation are accurately known.

A pH meter can normally measure with an accuracy of ±0.01 pH unit. Assuming that pH = -log[H^+], what is the concentration of hydrogen ion and its standard deviation in a solution whose pH is measured as 4.39?

$y = [H^+] = 10^{-4.39} = 4.074\text{x}10^{-5}$ mol/L

Since the mathematical operation involves taking an antilog to the base ten, Eqn. (1-17) applies.

$(s_y)_r = 0.01/0.43429 = 0.023$

However, the absolute standard deviation is required, so Eqn. (1-13) must be used to get the final result.

$s_y = (4.074\text{x}10^{-5})(0.023)$ mol/L $= 9.38\text{x}10^{-7}$ mol/L

The final result would be $[H^+] = (4.07\pm0.09)\text{x}10^{-5}$ mol/L.

What is the standard deviation of y in the calculation shown below?

$y = (1.375\pm0.003)(5.893\pm0.009) + 8.32\pm0.03$

In this example $y = 16.423$

This example is a bit more complex since both multiplication and addition are involved. Multiplication will involve relative standard deviations and addition will involve absolute standard deviations. Thus this type of problem must be done in steps, involving only one type of standard deviation per step. The order in which this is done will depend on the problem. In the above example the multiplication should be done first, followed by the addition. Care must be exercised not to intermix or confuse the different standard deviations in the various steps.

The multiplication step will be performed first, thus (1.375)(5.893) = 8.1029. The standard deviation of this intermediate result must be calculated using Eqn. (1-12), which requires relative standard deviations.

s_r of 1.375 = 0.003/1.375 = $2.18\text{x}10^{-3}$

s_r of 5.893 = 0.009/5.893 = $1.53\text{x}10^{-3}$

s_r of 8.1029 = $\{(2.18\text{x}10^{-3})^2 + (1.53\text{x}10^{-3})^2\}^{1/2} = 2.66\text{x}10^{-3}$

Now the summing step must be performed: thus 8.1029 + 8.32 = 16.423, the final result. The standard deviation of the final result must be calculated using Eqn. (1-11), which requires absolute standard deviations. The value for the number 8.32 is already known (0.03), but the value for 8.1029 must be calculated from its relative standard deviation using Eqn. (1-13). (These interconversion steps are easy to forget when doing this type of calculation.)

s of 8.1029 = $2.66\text{x}10^{-3}(8.1029) = 2.16\text{x}10^{-2}$

s of 16.423 = $\{(2.16\text{x}10^{-2})^2 + (0.03)^2\}^{1/2} = 0.037 = 0.04$

Thus the final result would be reported as 16.42±0.04.

In the above example, the use of Eqn. (1-14) might have saved a few steps in the calculation. However, when learning how to do these types of problems, the use of "shortcuts" may be

counterproductive. Almost all "real world" calculations involve a mixture of types of operations and thus involve both absolute and relative standard deviations. Successful application of the rules for propagation of errors requires attention to interconversions between the two types of standard deviations. The final example again illustrates this.

Assume that you are preparing a solution of NaCl. You weigh 0.2434 g of NaCl, place this in a 100 mL volumetric flask, and dilute to the mark with distilled water. You wish to estimate the standard deviation of the molarity of the resulting solution. In this example we will assume that the salt is 100.00% pure, that the volumetric flask holds exactly 100.00 mL of solution if properly filled, and that the balance is in perfect calibration.

The exact data are as follows: the GFM of NaCl is 58.443 g/mol, the mass of the weighing paper is 0.1374 g, that of the paper plus salt is 0.3808 g, the standard deviation of the GFM is ±0.001 g/mol, that of a single weighing is ±0.0002 g, and that of filling a 100 mL vol. flask is ±0.02 mL.

This may seem like a very difficult problem since you need an equation in order to apply the rules for propagation of error, and you have not been given one. Thus the first step is to derive an equation that presents the result as a function of all the variables. This is the first step in all such problems. In this case the first step is the derivation of the equation for the molarity, followed by calculation of the molarity.

M = moles/volume = {mass/GFM}/vol.

M = {(0.3808 g - 0.1374 g)/58.443 g/mol}/0.1000 L = 0.04165 mol/L

It can be seen from the above that one subtraction and two divisions are involved in the calculation. Thus you are again faced with the problem of separating the calculation into steps involving different operations. In this case the subtraction step (involving absolute standard deviations) will be performed before the division steps (involving relative standard deviations). Eqn. (1-11) is used to calculate the standard deviation of the mass, 0.2434 g.

$$s_{mass} = \{(0.0002\ g)^2 + (0.0002\ g)^2\}^{1/2} = 0.00028\ g$$

For the divisions Eqn. (1-12) is used and the result of both divisions can be calculated in a single step.

$$(s_M)_r = \{(s_{mass})_r^2 + (s_{GFM})_r^2 + (s_{vol})_r^2\}^{1/2}$$

However, this equation requires the use of relative standard deviations, and these must be calculated.

$$(s_{mass})_r = 0.00028\ g/0.2434\ g = 1.2\text{x}10^{-3}$$

$$(s_{GFM})_r = (0.001\ g/mol)/(58.443\ g/mol) = 1.7\text{x}10^{-5}$$

$$(s_{vol})_r = 0.02\ mL/100\ mL = 2.0\text{x}10^{-4}$$

The semifinal calculation can now be performed.

$$(s_M)_r = \{(1.2\text{x}10^{-3})^2 + (1.7\text{x}10^{-5})^2 + (2.0\text{x}10^{-4})^2\}^{1/2}$$

$$(s_M)_r = \{1.4\text{x}10^{-6} + 2.9\text{x}10^{-10} + 4.0\text{x}10^{-8}\}^{1/2} = 1.2\text{x}10^{-3}$$

Since the result of the above calculation is a relative standard deviation and the absolute deviation is usually desired, the final calculation must use Eqn. (1-13) to achieve the desired result.

$s_M = M(s_M)_r = (0.04165 \text{ mol/L})(1.2\text{x}10^{-3}) = 5.0\text{x}10^{-5}$ mol/L

The final result would thus be reported as (0.04165±0.00005) M

Note in the above example that the uncertainty in the GFM had no effect on the final result and it could have been ignored in the calculation (as is usually the case). The step involving dilution in the volumetric flask had such a small standard deviation that it did not affect the final result either. The only step that contributed significantly to the uncertainty in the answer was the weighing. If a more precise result were desired, the weighing step would be the first to be examined for improvement.

Also regarding the above example, the question might be asked - what effect would it have if the salt was only 99.8% pure, or the volumetric flask held 100.2 mL when properly filled, or the balance weighed light by 0.0003 g? These are all determinate errors which would affect the accuracy of the final result (the molarity of the solution), but they would not affect the calculated standard deviation, since they are constant errors and do not affect the precision of the steps in the procedure in which they are involved. In the absence of determinate error, the standard deviation gives an estimate of the error in the result (or the accuracy of the result). If determinate error is present, the standard deviation indicates the precision of a measurement, but no longer provides an estimate of the accuracy. This situation would be analogous to that depicted in Figure 1-2, Case III.

1.3 LINEAR LEAST SQUARES CURVE FITTING

In science the behaviours of many systems follow mathematical relationships, often linear relationships. The goal of many investigations is to discover these relationships. This is normally done by measuring a dependent variable, y, as a function of one or more independent variables, x_i, and plotting these data appropriately with the mathematical relationship which is believed to apply to the system in mind. If the plot exhibits the expected behaviour, *e.g.* a straight line, then the relationship is confirmed. When measuring a single quantity, the best result is obtained by making a set of measurements and taking the mean rather than relying on a single measurement. The reason for this, as explained above, is that errors will enter into the procedure to provide a distribution of values rather than a single value. The same effect will occur when making measurements at more than one value of x; errors will occur which produce scatter in the data points for the plot. The question then arises - is there some method of obtaining the best possible curve through data points with scatter?

Least squares curve fitting, more properly called least squares regression analysis, is a method of obtaining the "best" fit of experimental data to a model (a mathematical equation). The best known example of this is fitting data to a straight line, $y = mx + b$. This is especially easily done with computers and many computer programs, including all spreadsheet programs, contain procedures to perform this operation. Since many relationships in science are believed to be linear and since the fitting is done by computer, there is an unfortunate tendency to accept the results of such fittings without question. Thus some understanding of the least squares method should be made available to those who use the method. That is the purpose of this section.

1.3.1 THE GENERAL METHOD OF LEAST SQUARES

The general mathematical model for the behaviour of a system can be written as:

$$y = f(x_1, x_2, \cdots, x_m, a_1, a_2, \cdots, a_k) \qquad (1\text{-}18)$$

where y is the dependent variable, the x_j are the independent variables, and the a_k are constants that define the model, often called parameters.

For a linear, first order case with one independent variable (the most familiar situation)

$$y = a_1 x + a_0 \tag{1-19}$$

Errors in measurement will provide enough scatter in the measured values of the variables to prevent the measured data points from falling on the curve to which the data are being fit. The residuals are defined as the differences between the measured values of x_{ij} and y_i and those calculated from the model equation.

$$(R_y)_i = (Y_i - y_i) \tag{1-20}$$

$$(R_x)_{ij} = (X_{ij} - x_{ij}) \tag{1-21}$$

where Y_i is the ith experimental value of y, y_i is the value calculated from the model equation, X_{ij} is the ith experimental value for the jth independent variable, x_j, and x_{ij} is the value calculated from the model equation. For the simple first order linear case above, j = 1 (there is only one independent variable). These ideas are illustrated, also for the first order linear case, in Figure 1-4.

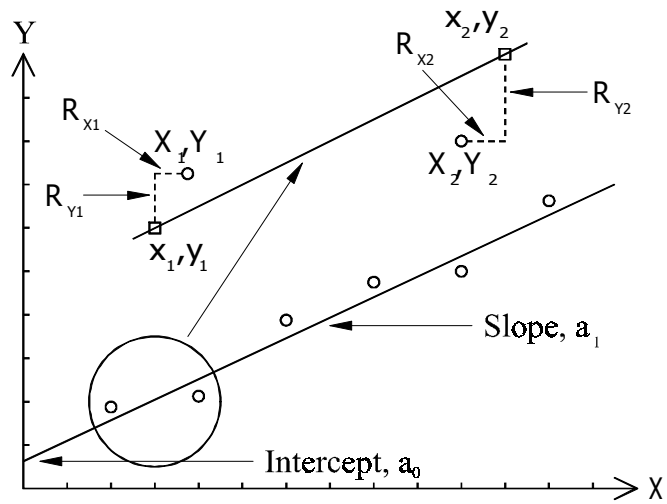

Figure 1-4

The sum of the squares of the residuals (RSS) is now defined as

$$\text{RSS} = \sum_{i=1}^{n} \left[(R_y)_i^{\,2} + \sum_{j=1}^{m} (R_x)_{ij}^{\,2} \right] \tag{1-22}$$

where n is the number of data point sets in the data and m is the number of independent variables. The residuals, and thus the sum of the squares of the residuals, change as the parameters (the a_k) change, since changing the parameters changes the position of the model curve but not the experimental data.

The "best" values of the model parameters are defined as those that produce a minimum in RSS. This forces the plot to approach all the data point sets as closely "on average" as possible.

The First Order Linear Case

We could continue with the general treatment; however, the most useful case at the level of this text is the first order linear model where m = j =1. Thus this model ($y = a_1 x + a_0$) will be treated in some detail. If m = j = 1, then the sum of the squares of the residuals becomes

$$\text{RSS} = \sum_{i=1}^{n} \left[(R_y)_i^{\,2} + (R_x)_i^{\,2} \right] \tag{1-23}$$

It is a difficult task to properly estimate the amount of error to assign to y_i and to x_i for any data point pair. In other words it is difficult to simultaneously assign values to both $(R_y)_i$ and $(R_x)_i$. Thus a simplifying assumption is now almost always made, and it is very important to understand this assumption. In many cases the errors in x are much less than those in y. In other words X can be set much more precisely than Y can be measured. If this is so, then the problem simplifies greatly. At his point we will make that assumption.

$$(R_x)_i \langle\langle (R_y)_i \tag{1-24}$$

and

$$\text{RSS} = \sum (R_y)_i^2 \tag{1-25}$$

It can now be stated that

$$(R_y)_i = (Y_i - y_i) = (Y_i - a_1 X_i - a_0) \tag{1-26}$$

since $x_i = X_i$. The symbol Σ is taken to represent the sum from i = 1 to i = n. Since there is assumed to be no error in x, the experimental points will lie directly above or below the corresponding fitted points on the model line and the residuals as depicted in Figure 1-4 now change to those shown in Figure 1-5. The expression for the sum of squares of the residuals now becomes

$$\text{RSS} = \sum \left[Y_i - a_1 X_i - a_0\right]^2 \tag{1-27}$$

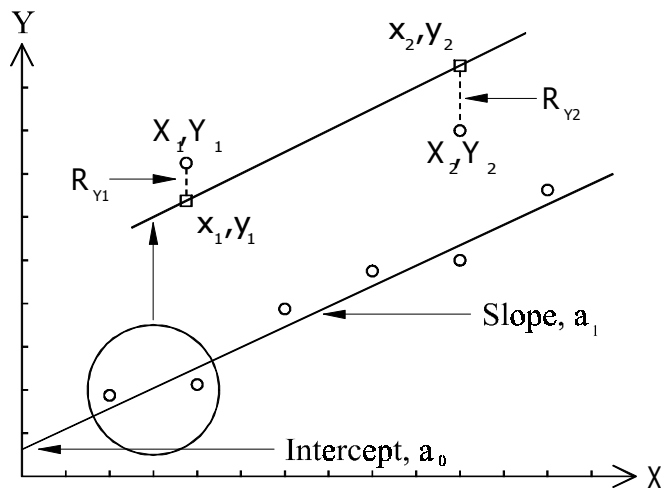

Figure 1-5

We now need to determine which values of a_1 and a_0 minimize RSS, these will be the "best" values of a_1 and a_0. This is done by taking the first derivative of RSS with respect to both a_1 and a_0, and setting these equal to zero.

$$\frac{\partial \text{RSS}}{\partial a_1} = 0 = \sum \left[2(Y_1 - a_1 X_i - a_0)(-X_i)\right] \tag{1-28}$$

or

$$0 = \sum \left[-X_i Y_i + a_1 X_i^2 + a_0 X^i\right] \tag{1-29}$$

or

$$a_1 \sum X_i^2 + a_0 \sum X_i = \sum X_i Y_i \tag{1-30}$$

and

$$\frac{\partial \text{RSS}}{\partial a_0} = 0 = \sum \left[2(-1)(Y_i - a_1 X_i - a_0)\right] \tag{1-31}$$

or

$$0 = \sum \left[-Y_i + a_1 X_i + a_0\right] \tag{1-32}$$

or

$$a_1 \sum X_i + a_0 n = \sum Y_i \tag{1-33}$$

Eqns. (1-30) and (1-33) are two simultaneous linear equations which are easily solved by standard methods and provide the least squares or "best" values for a_1 and a_0. The results of this operation are as follows:

$$a_1 = \frac{\sum X_i \sum Y_i - n \sum X_i Y_i}{\left(\sum X_i\right)^2 - n \sum X_i^2} \tag{1-34}$$

$$a_0 = \frac{\sum X_i \sum X_i Y_i - \sum X_i^2 \sum Y_i}{\left(\sum X_i\right)^2 - n \sum X_i^2} \tag{1-35}$$

One calculates all the sums (ΣX_i, ΣX_i^2, ΣY_i, $\Sigma X_i Y_i$), plugs these into Eqns. (1-34) and (1-35), and calculates a_1 and a_0. Slightly different, but equivalent, expressions (algorithms) are often found in the literature. Two features of these equations should be noted, (a) the denominators in the two expressions are the same and thus only need to be calculated once, and (b) $(\Sigma X_i)^2 \neq \Sigma X_i^2$.

As an example, consider the *X-Y* data set shown below in the first two columns, which is assumed to be first order (this data set will be revisited in Chapter 3). Columns 3 and 4 are needed for the sums in the least squares calculations and have been calculated from columns 1 and 2.

	X	*Y*	X^2	*XY*
	0.000	0.000	0.0	0.0
	$2.00\text{x}10^{-5}$	0.189	$4.00\text{x}10^{-10}$	$3.780\text{x}10^{-6}$
	$4.00\text{x}10^{-5}$	0.383	$1.60\text{x}10^{-9}$	$1.532\text{x}10^{-5}$
	$6.00\text{x}10^{-5}$	0.571	$3.60\text{x}10^{-9}$	$3.426\text{x}10^{-5}$
	$8.00\text{x}10^{-5}$	0.756	$6.40\text{x}10^{-9}$	$6.048\text{x}10^{-5}$
	$1.00\text{x}10^{-4}$	0.948	$1.00\text{x}10^{-8}$	$9.480\text{x}10^{-5}$
SUM	$3.00\text{x}10^{-4}$	2.847	$2.20\text{x}10^{-8}$	$2.0864\text{x}10^{-4}$

Thus $\Sigma X_i = 3.00\text{x}10^{-4}$, $\Sigma Y_i = 2.847$, $\Sigma X_i^2 = 2.20\text{x}10^{-8}$, $\Sigma X_i Y_i = 2.0864\text{x}10^{-4}$, $(\Sigma X_i)^2 = 9.00\text{x}10^{-8}$, n = 6

$a_0 = \{3.00\text{x}10^{-4}(2.0846\text{x}10^{-4}) - 2.20\text{x}10^{-8}(2.847)\}/\{9.00\text{x}10^{-8} - 6(2.20\text{x}10^{-8})\}$

$a_0 = 4.2\text{x}10^{-11}/4.2\text{x}10^{-8} = 0.001$

$a_1 = \{3.00\text{x}10^{-4}(2.847) - 6(2.0864\text{x}10^{-4})\}/-4.20\text{x}10^{-8} = 9.47\text{x}10^{3}$

The best straight line through the data in columns 1 and 2 will thus be

$Y = 9.47\text{x}10^{3}X + 0.001$

Note that including the first point (*X*=0, *Y*=0) does not force the intercept to be zero. The question is often asked, "should such a point be included in the data set"? The answer is *YES*, as long as it is known that *Y* is actually zero when *X* is zero.

Individuals seldom need to use the above equations to obtain the best values for the parameters of a first order linear model, since there are many computer programs written to provide this information, including all spreadsheet programs and statistics packages. These programs may also provide additional information such as estimates of the standard deviations of the parameters. When using such programs it should be kept in mind that it is assumed that: (a) all the error is found in the *y* values and (b) that all data have the same precision. If either or both of these assumptions are incorrect, more sophisticated fitting methods are available.

A word of caution is in order - the method of least squares can fit any data to a linear model (first order equation), the data do not have to make a straight line. Furthermore, the results do not necessarily indicate whether the data are truly linear. Thus it is always advisable to plot the data along with the resulting least squares line to determine whether a linear model should be applied. Observation of the resulting graph is an excellent way to check the "goodness of fit".

1.3.2 SOME GENERAL CONSIDERATIONS

A very specific, but very useful, case, the linear first order model, has been treated in some detail:

$$y = a_1x + a_0, \text{ all error in } y\text{, all values of } y \text{ have equal variance.}$$

What will occur if the model is more complicated? One might imagine several examples:

1. The order of the equation is greater than one, *e.g.* $y = a_0 + a_1x + a_2x^2 + \ldots$
2. There is more than one independent variable, *e.g.* $y = a_0 + a_1x_1 + a_2x_2 + \ldots$
3. Some combination of 1. and 2., *e.g.* $y = a_0 + a_1x_1 + a_2x_1^2 + a_3x_2 + a_4x_2^2 + a_5x_1x_2 + \ldots$

If in the above all error can be assumed to be in the y values, then all of these examples fall into the "linear" least squares category and exact solutions are available using the same mathematical procedure as above (the algebra becomes somewhat more complicated and the methods will not be discussed here). What, then, is meant by linear least squares? A linear least squares solution is available when y is linear in the parameters, not in the independent variables. Thus a second or higher order polynomial is a linear case, since y is linear in the a_k.

What happens if either (or both):

1. y is not linear in the parameters, *e.g.* $y = a_0 + a_1\sin(a_2x)$. Here y is linear in a_0 and a_1, but not a_2.
2. The error in x is approximately equal to or greater than that in y.

Now the system becomes non-linear. One approach is to use a Taylor expansion of $f(x_j, a_k)$ and neglect all terms higher than first order. In other words, one linearizes the system. Guesses must now be made for the values of all the parameters and an iterative improvement process is used to zero in on the "correct" values for the parameters. The success of this process depends, in part, on the accuracy of the initial guesses. The results are not exact, but are usually the best that can be accomplished. Computer programs are also available to do this sort of least squares fitting.

1.4 QUANTITATIVE CALCULATIONS INVOLVING CONCENTRATION

Before discussing other types of quantitative calculations, a review and discussion of various methods of expressing concentration is in order.

1.4.1 METHODS FOR EXPRESSING CONCENTRATION.

There are a number of different methods for expressing the concentration of a solute in a solution. Only those of importance in analytical chemistry will be discussed below.

Weight by weight methods:

Three of the several methods are presented here. All weight by weight methods are very similar.

1) Percent by weight or parts per hundred. This method is usually used for concentrated reagents such as sulfuric acid or ammonia. Percent by weight is defined simply as the grams of solute in 100 grams of *solution* (not solvent). More specifically:

$$\% \text{ by wt.} = (\text{g solute/g solution})(100) \qquad (1\text{-}36)$$

Two examples will be given to illustrate this method of specifying concentration.

If 1.00 g of NaCl is dissolved in 10.0 g of H_2O, the % by wt. of NaCl is [1.00 g/(1.00 g+10.0 g)](100) or 9.09%.

Reagent sulfuric acid is 96% by wt. This means that out of every 100 g of solution 96 g are H_2SO_4 and 4 g are H_2O.

Compute the volume of 96.0% H_2SO_4 needed to prepare 250 mL of a 1.00 molar solution, if the density of the reagent is 1.81 g/mL.

Step 1: calculate the moles of pure H_2SO_4 needed. From the definition of molarity (see below)
moles = (volume)(molarity) = (0.250 L)(1.00 mol/L) = 0.250 mol.

Step 2: calculate the mass of pure H_2SO_4 needed.
mass = (mol)(GFM) = (0.250 mol)(98.08 g/mol) = 24.52 g.

Step 3: calculate the mass of 96% reagent needed.
96.0% of the mass of reagent = 24.52 g, or

(0.960)(mass of reagent) = 24.52 g.

Thus mass of reagent = 24.53 g/0.960 = 25.54 g.

Step 4: calculate the volume of 96% reagent needed. From the definition of density, d = mass/volume, one obtains

volume = mass/d = 25.54 g/1.81 g/mL = 14.1 mL

2) Parts per million or ppm. This method is used for dilute solutions. It is defined simply as the grams of solute in 10^6 g of solution.

$$\text{ppm} = (\text{g solute/g solution})(10^6) \qquad (1\text{-}37)$$

If 2.5 mg of NaCl is dissolved in 1000 g of H_2O, 1.0 mg of Na^+ will be present.
Thus ppm Na = $\{0.0010\text{g}/(1000\text{g}+0.0025\text{g})\}(10^6)$ = 1.0 ppm.

If the solution is very dilute (often the case for such solutions), the density can be taken as 1.00 g/mL or 1000 g/L and 1.00 L will weigh 1000 g. In this case the definition of ppm becomes

$$\text{ppm} \approx (\text{mg of solute})/(\text{litres of solution}) \qquad (1\text{-}38)$$

NOTE : this definition cannot be used for nonaqueous solutions since such solutions rarely have a density of 1.00 g/mL.

How much $Fe(NH_4)_2(SO_4)_2{\cdot}6H_2O$ (ferrous ammonium sulfate, GFM = 392.16 g/mol) must be weighed out to prepare 250 mL of a 10.0 ppm solution of Fe(II) (GAM of Fe = 55.847 g/mol)? A 10.0 ppm solution contains 10.0 mg/L of Fe(II) (*NOT 10.0 mg/L of ferrous ammonium sulfate*).

mass of Fe = (10.0 mg/L)(0.250 L) = 2.50 mg

mass of $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ = (2.50 mg)(392.16 g/mol)/(55.847 g/mol) = 17.6 mg.

---

3) Parts per billion or ppb. This method is used for *extremely* dilute solutions. It is defined as grams of solute in 10^9 g of solution.

$$\text{ppb} = (\text{g solute/g solution})(10^9) \qquad (1\text{-}39)$$

For solutions that have a density of approximately 1000 g/L

$$\text{ppb} \approx (\mu\text{g of solute})/(\text{litres of solution}) \qquad (1\text{-}40)$$

A solution which contains 5.0 μg of NaCl in 500 mL of solution will contain 2.0 μg of Na^+. Such a solution will contain (2.0 μg Na^+)/0.500 L or 4.0 ppb Na.

Formality.

This is a method of expressing concentration with which you are probably unfamiliar. It is used to differentiate that which is dissolved in solution from that which actually exists in solution at equilibrium. Some chemists refer to formality as "analytical concentration" because it is the numerical value determined in most analyses. The formality of a solution is the number of formula weights or moles of a substance dissolved in or placed in a litre of *solution*. This does not consider what happens to that substance after it is in the solution. In other words it does not take account of any chemical reactions that might occur in solution. Specifically

$$\text{Formality} = F = (\text{mol solute } \textit{dissolved})/(\text{volume of soln.}) = \text{mol}/V \qquad (1\text{-}41)$$

The units or dimensions are mol/L. Some analytical texts do not use formality (most mention it) but it will be used in this text, especially when discussing acid-base equilibria.

Molarity.

Molarity is the number of moles of a substance *actually present* per litre of solution, or

$$\text{Molarity} = M = (\text{mol solute } \textit{present})/(\text{volume of soln.}) = \text{mol}/V \qquad (1\text{-}42)$$

The units of molarity are mol/L. Unfortunately these are the same units as formality and often molarity and formality are confused. Remember that formality represents what is placed in solution, molarity represents what actually exists in solution at a given time. The molarity of a material is often represented by its chemical symbol or formula written within square brackets, *e.g.* $[Na^+]$ (this indicates the actual material present in solution). Formalities are *never* written this way.

If one dissolves 58.5 g of NaCl (GFM of NaCl = 58.5 g/mol) in enough water to form 1.00 L of solution, the formality of NaCl will be 1.00 mol/L. However, NaCl does not exist as such in solution; it completely dissociates to form Na^+ and Cl^- ions. 1.00 mol of NaCl will form 1.00 mol of both Na^+ and Cl^- ions. Thus one does not speak of the molarity of NaCl, but rather of the molarity of Na^+ and Cl^-. $[Na^+] = [Cl^-] = 1.00$ mol/L.

6.005 g of pure acetic acid (GFM = 60.05 g/mol) are dissolved in water to form 1.000 L of solution. What are the formalities and molarities of all the solution components? Assume that acetic acid ionizes 3.0 % in this solution.

Mol acetic acid *dissolved* = 6.005 g/60.05 g/mol = 0.1000 mol

F = formality = 0.1000 mol/1.000 L = 0.1000 mol/L

For the purposes of this problem it will be assumed that acetic acid reacts as follows when dissolved in water.

$$CH_3COOH \rightarrow CH_3COO^- + H^+$$

If 3.0% of the acid ionizes, it can be stated that

mol CH_3COO^- = mol H^+ = (0.030)(0.1000 mol) = 0.0030 mol

mol unionized CH_3COOH = 0.1000 mol - 0.0030 mol = 0.0970 mol

$[H^+] = [CH_3COO^-]$ = 0.0030 mol/1.000 L = 0.0030 M

$[CH_3COOH]$ = 0.0970 mol/1.000 L = 0.0970 M

What is the molarity of a dilute solution (d = 1.00 g/mL) of 0.10 ppm Cu^{2+} (GAM = 63.546 g/mol)? If the solution is dilute, the definition of ppm can be taken as ppm = mg/litre.

0.10 ppm = 0.10 mg/L = 1.0×10^{-4} g/L

To go from g/L to mol/L one needs to calculate the number of moles in 1.0×10^{-4}g.

mol = 1.0×10^{-4} g/63.546 g/mol = 1.57×10^{-6} mol.

Thus 0.10 ppm Cu(II) = 1.57×10^{-6} mol/L = 1.6×10^{-6} M

1.4.2 CALCULATIONS INVOLVING DILUTION.

Often samples subjected to analytical procedures are diluted before analysis, yet the results are needed for the original samples, not the diluted solutions. Also many very dilute solutions are prepared by a technique called serial dilution since the amounts of solute required are too small to weigh directly. In both cases calculations involving dilution are required and should now be discussed.

The important thing to remember when performing calculations involving dilutions is that the number of moles of material is the same in the solution both before and after dilution. This fact can always be used to determine the concentration either before or after dilution.

5.00 mL of a 0.0400 F solution of HCl are diluted to 100 mL. What is the final concentration of HCl?

Since F = mol/V (by definition), then mol = F*V*

mol HCl (in both solutions) = (0.0400 mol/L)(0.00500 L) = 2.00×10^{-4} mol.

Formality (of the second solution) = mol/*V* = 2.00×10^{-4} mol/0.100 L = 0.00200 F.

What volume of a 1.00 F solution of ammonia (NH_3) is needed to prepare 500 mL of a 0.0250 F solution?

mol NH_3 (both before and after dilution) = (0.0250 mol/L)(0.500 L) = 0.0125 mol.

V of 1.00 F NH_3 = mol/F = 0.0125 mol/1.00 mol/L = 0.0125 L = 12.5 mL.

10.0 mL of a sample are diluted to 50.0 mL before analysis in order to add reagents for the analysis. The concentration of Ca^{2+} in the diluted solution is found to be 0.00354 M. What is the concentration of calcium in the original sample?

mol Ca^{2+} (in both solutions) = (0.00354 mol/L)(0.0500 L) = 1.77×10^{-4} mol.

Molarity of Ca^{2+} in original sample = 1.77×10^{-4} mol/0.0100 L = 0.0177 M.

Often such calculations can be simplified by the use of dilution factors. These factors are the ratios of the initial volume to the final volume or *vice versa* depending on whether the final or initial concentration is to be calculated. Thus

$$C_f = C_i\left(V_i / V_f\right) \tag{1-43}$$

and

$$C_i = C_f\left(V_f / V_i\right) \tag{1-44}$$

where C is concentration, V is volume, i refers to initial conditions, and f refers to final conditions. Equations (1-43) and (1-44) come directly from the fact that the moles are the same before and after dilution, $C_iV_i = C_fV_f$. To remember which volume goes in the numerator recall that the final concentration will always be less than the initial concentration. Thus when calculating final concentration, the smaller volume goes in the numerator of the dilution factor and *vice versa*. In the cases of the first and last examples given above.

Formality of HCl = (0.0400 F)(5.00 mL/100 mL) = 0.00200 F

Molarity of Ca^{2+} = (0.00354 M)(50.0 mL/10.0 mL) = 0.0177 M

If you are unsure about the use of dilution factors, use the first method involving moles. Although it is longer, it will produce correct results without guessing about which way to set up the ratio of the dilution factor.

A second area in which dilution is often involved is in the preparation of solutions for calibration curves (refer to Chapter 3). An example will best illustrate this type of calculation.

You wish to prepare at least 50 mL each of solutions containing 0.50, 1.00, 2.00, and 4.00 ppm Pb^{2+} for the calibration curve in a lead analysis. The most concentrated of these solutions will contain only 4.00 mg/L of lead. An ordinary analytical balance can weigh accurately to only about ±0.2 mg. To obtain a precision of about 1 in 400 (0.01 mg in 4.00 mg) at least 80.0 mg of lead must be weighed. Assuming that you use $Pb(NO_3)_2$ (GFM = 331.2 g/mol) as your source of lead (GAM = 207.2 g/mol), you must weigh

mass of lead nitrate = (80.0 mg)(331.2 g/mol/207.2 g/mol) = 127.9 mg

If this amount of material is dissolved in 1.00 L of water, then the concentration of Pb(II) will be 80.0 ppm (80.0 mg/L). This must be diluted by a factor of 20 to make the 4.00 ppm solution desired (80.0 ppm/20 = 4.00 ppm). One way in which this can be accomplished is by diluting 5.00 mL of the 80.0 ppm solution to 100 mL. It could also be done by diluting 25.0 mL to 500 mL - the dilution factor in both cases is 20 (100 mL/5 mL = 500 mL/25 mL = 20).

Conc. Pb(II) = (80.0 ppm)(5.00 mL/100 mL) = 4.00 ppm

The 2.00 ppm solution can be prepared by diluting the 4.00 ppm solution, requiring a dilution factor of two (4.00 ppm/2.00 ppm = 2). This could be achieved by diluting 50.0 mL of the 4.00 ppm solution to 100.0 mL, leaving 50 mL of the 4.00 ppm solution for the calibration curve (assuming it is not needed for any other purpose).

Conc. Pb(II) = (4.00 ppm)(50.0 mL/100 mL) = 2.00 ppm

Likewise the 1.00 ppm solution can be prepared from the 2.00 ppm solution by a factor of two dilution, and the 0.50 ppm solution can be prepared from the 1.00 ppm solution in a similar fashion. The final result would be 50.0 mL each of 4.00, 2.00, and 1.00 ppm and 100.0 mL of 0.50 ppm for the calibration curve.

If 500.0 mL of the 4.00 ppm solution had been prepared (by diluting 25 mL of the 80 ppm solution to 500 mL) instead of 100.0 mL, the 2.00 ppm could have been prepared in the same fashion (by diluting 50.0 mL of the 4.00 ppm solution to 100.0 mL). However, the 1.00 ppm solution could have been prepared from the 4.00 ppm solution using a dilution factor of 4 (dilute 25.0 mL of the 4.00 ppm solution to 100.0 mL). The 0.50 ppm solution could have been prepared from the 2.00 ppm solution also using a dilution factor of 4 (by diluting 25.0 mL of 2.00 ppm to 100.0 mL). The final result would be 425 mL of 4.00 ppm, 75 mL of 2.00 ppm, and 100 mL each of the 1.00 and 0.50 ppm solutions.

This illustrates that there are usually several ways to prepare the same series of serially diluted solutions. The method chosen will depend on several factors including the concentrations desired, the volumes of solutions needed, and the volumetric glassware available. The first scheme explained above would require a 5.00 mL pipette, 25 mL (or 50.0 mL) pipettes and 100.0 mL volumetric flasks and would provide exactly the amounts of solutions needed. The second scheme would require 25 mL pipettes, one 500.0 mL and several 100.0 mL volumetric flasks. It would provide extra solution volumes in case of additional requirements. There is no general formula that can be applied to such calculations, each case is unique.

1.5 QUANTITATIVE CALCULATIONS INVOLVING TITRATIONS

The titration is one of the oldest methods of performing quantitative analyses, yet it is still extremely useful. Titrations have two very important features: they employ very simple and inexpensive "instrumentation", the buret, and they are generally one of the most accurate methods of analysis available. The accuracy often achievable is 0.1% and it can sometimes be extended to 0.01%. However, titrations are usually not capable of trace analysis, that is analysis of components which make up less than 1% of a sample or components that are less than about 10^{-3} M in solution concentration. Thus titrations are used for analysis of major sample components or fairly concentrated solutions when high accuracy is required. They are also used when only a few samples need analysis since it is usually easier to set up a titration than a complicated instrumental procedure.

Most readers will be familiar with titrations. They are nothing more than chemical reactions that are carried out under strictly controlled conditions with care taken to determine exactly when the reaction is finished. This concept leads to two definitions that should be stressed at this time.

1) The *equivalence point* is that point in the titration at which exactly enough titrant (the reagent in the buret) has been added to completely consume all the titrand (the reagent in the reaction flask or beaker) "in theory".

2) The *end point* is that point at which the indicating device or substance signals the equivalence point. It is the point the analyst takes as the equivalence point.

One wants the end point to equal the equivalence point (or else there is an error) and for the present this will be considered to be the case. Titrations will be discussed in general because all titrations are very similar - if you thoroughly understand one, you understand them all.

It is customary to place the reactant whose concentration is known (the standard solution) in the buret, thus making it the *titrant*. The sample which contains the *analyte*, the substance whose concentration is to be determined, is placed in the reaction vessel, making it the *titrand*. This is usually the most convenient way to carry out a titration. In the following discussions T will represent the titrant, U will represent the unknown (the analyte) or titrand, and P_1, P_2, *etc*. will represent the products of the titration reaction .

1.5.1 DIRECT TITRATIONS

In all titrations the calculations depend on the stoichiometry of the titration reaction, to be specific, the relationship between the moles of T and U in the reaction. The moles of P_1 and P_2 do not enter into the calculation. The following general reaction will be assumed

$$aT + bU \rightarrow cP_1 + dP_2 + \ldots$$

At the equivalence point exactly enough T has been added to just react with all the U and there is "no" T or U in the titration flask (assuming that the titration reaction goes completely to the right). Thus it can be stated that

$$\text{(original mol U)/(mol T added)} = b/a \qquad (1\text{-}45)$$

Since the moles of U are usually the quantity of interest, this is the quantity to be determined.

$$\text{original mol U} = (b/a)(\text{mol T added}) \qquad (1\text{-}46)$$

Note that the products do not enter the calculation, but that the reaction stoichiometry does through the ratio (b/a). Thus one ***must*** have a balanced equation to do such calculations. The moles of titrant are calculated from the formality of the titrant, F_T. This is done using the basic relation between moles, formality, and volume.

$$\text{mol} = (\text{formality})(\text{volume}) = FV \qquad (1\text{-}47)$$

A set of equations for direct titrations can now be derived for different situations. Let V_T represent the volume of titrant added at the equivalence point.

1) As indicated above, the first step is to solve for the moles of U.

$$\text{mol U} = (b/a)(F_T V_T) \qquad (1\text{-}48)$$

Normally one wishes to know more than just the moles of U present in the titration flask.

2) If the sample is a liquid, one often needs to know the formality of the analyte in the liquid.

$$\text{mol U} = F_U V_U, \tag{1-49}$$

where V_U is the ***original*** volume of the sample (**NOT the volume in the titration vessel at the equivalence point.**) Thus

$$F_U V_U = (b/a)(F_T V_T) \tag{1-50}$$

or

$$F_U = (b/a)(F_T V_T)/V_U \tag{1-51}$$

Note in this case the original volume of sample must also be accurately known.

3) If the sample is a solid which has been dissolved in some solvent for the titration, then the mass of analyte may be the desired quantity.

$$\text{mass U} = (\text{mol U})(\text{GFM}_U) \tag{1-52}$$

or

$$\text{mass U} = (\text{GFM}_U)(b/a)(F_T V_T) \tag{1-53}$$

Note that, in this case, the sample volume does not enter into the calculation - it does not matter.

4) However, for a solid one generally wants to know the percentage of analyte in the sample.

$$\text{\%U in sample} = (\text{mass U})(100)/(\text{sample mass}) \tag{1-54}$$

or

$$\text{\%U} = (\text{GFM}_U)(b/a)(F_T V_T)(100)/(\text{sample mass}) \tag{1-55}$$

Practical aspects

The parameters above which must be accurately known for a successful titration are: b/a, which is determined from the chemical reaction, V_T (from the buret), V_U (usually from a pipet, if the sample is a liquid), GFM_U (from the periodic table), sample mass (from a balance), and F_T. The question arises, how does one know F_T accurately? F_T is usually determined in one of two different ways:

1) Make up the titrant carefully from a primary standard, *e.g.* $K_2Cr_2O_7$.

2) Make up the titrant from a material the purity of which is less than that of a primary standard, *e.g.* NaOH, and titrate against a primary standard, *e.g.* potassium biphthalate or potassium hydrogen phthalate (KHP - see Figure 1-6). This procedure is called standardization.

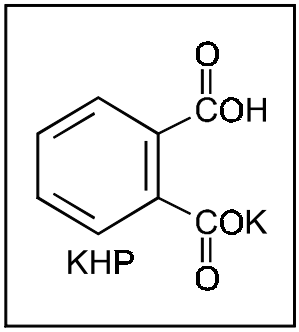

Figure 1-6

This of course raises another question - what is a primary standard? A primary standard is a compound that has the following characteristics:

1) an exactly known composition (*e.g.* known waters of hydration),

2) "no" impurities - at least 99.9% pure,

3) does not react with or take up the components of the atmosphere, *e.g.* CO_2, O_2, or H_2O,

4) has a form that is easily weighed (*e.g.* not a fine powder).

The quantity of primary standard involved in the analysis is determined by a weighing. Thus the final accuracy in almost all analytical procedures is a function of a weighing. The balance is the most basic analytical tool.

The same equations are used in a standardization titration as in the titration of an unknown. An example will demonstrate this.

Sulfuric acid is standardized by titration with primary standard Na_2CO_3. 0.5000 g of dried sodium carbonate requires 45.00 mL of H_2SO_4 in the standardization titration. The sulfuric acid is then used to determine the concentration of NH_3 in a sample solution. 50.00 mL of the sample solution require 33.00 mL of the sulfuric acid in the titration of the sample. What is the formality of ammonia in the sample?

a) Standardization:

1) Write the balanced chemical equation (*always do this*).

$$H_2SO_4 + Na_2CO_3 \rightarrow Na_2SO_4 + CO_2 + H_2O$$

2) Establish the relation between the moles of titrant and titrand. This is 1:1 in this case, thus

mol H_2SO_4 = mol Na_2CO_3.

3) Substitute the other relations for moles of titrant and titrand.

mol Na_2CO_3 = mass Na_2CO_3/GFM = 0.5000 g/106.0 g/mol

mol H_2SO_4 = $F_T V_T$ = (0.04500 L)F_T

(0.04500 L)F_T = 0.5000 g/106.0 g/mol

4) Solve for F_T

F_T = 0.5000 g/(0.04500 L)(106.0 g/mol) = 0.1048 mol/L

b) Titration of the sample:

1) Write the balanced equation.

$$H_2SO_4 + 2NH_3 \rightarrow (NH_4)_2SO_4$$

2) Establish the relation between the moles of titrant and titrand. This is 1:2 in this case, thus

mol NH_3 = 2(mol H_2SO_4)

3) Substitute the other relations for moles.

mol NH_3 = $F_U V_U$ and mol H_2SO_4 = $F_T V_T$

$F_U V_U = 2F_T V_T$

4) Solve for the formality of the analyte (NH_3).

$F_U = 2F_T V_T / V_U$ = 2(0.1048 mol/L)(0.03300 L)/(0.05000 L) = 0.1384 mol/L

It is recommended that the reader not try to memorize all the equations derived earlier, but rather try to solve the problem in a way similar to that presented in the example above.

1.5.2 BACK TITRATIONS

These are titrations in which a *known excess* of a standard material or solution is added to a sample. All the analyte reacts and some standard remains. The remaining standard is then titrated with a second standard, the titrant. From the known amount of the first standard and the volume and concentration of titrant, the amount or concentration of analyte is calculated. Back titrations are more complicated than simple titrations, involving more manipulations and more standard solutions. Thus they are used only when a simple titration will not work.

In the following "S" will represent the first standard solution, the one added in excess. All other symbols will retain the same meaning as above.

1) Excess standard solution is added to the sample, consuming all U.

$$aS + bU \rightarrow P_1 + P_2$$

$$\text{mol U} = (b/a)(\text{mol of S consumed}) = (b/a)(\text{mol } S_U) \qquad (1\text{-}56)$$

2) Some S, S_{xs}, remains unused. The S_{xs} is then titrated with T

$$cT + dS \rightarrow P_3 + P_4$$

$$\text{mol } S_{xs} = (d/c)(\text{mol T at eq. pt.}) \qquad (1\text{-}57)$$

$$\text{Total mol S} = \text{mol } S_T = \text{mol } S_{xs} + \text{mol } S_U \qquad (1\text{-}58)$$

$$\text{mol } S_T = (d/c)(\text{mol T}) + (a/b)(\text{mol U}) \qquad (1\text{-}59)$$

3) Solve equation (1-59) for the moles of U.

$$\text{mol U} = \{\text{mol } S_T - (d/c)(\text{mol T})\}(b/a) \qquad (1\text{-}60)$$

or

$$\text{mol U} = (b/a)\{V_S F_S - (d/c)(V_T F_T)\} \qquad (1\text{-}61)$$

if S is a solution.

One can proceed as was done above to obtain expressions for the mass of U, F_U or %U. This will not be repeated here. Instead an example of a back titration will be presented.

A 1.000 g sample containing $CaCO_3$, a material insoluble in pure water, is dissolved in 100.0 mL of 0.1000 F HCl (an excess) and boiled to dispel the CO_2 formed. The remaining HCl is titrated with 0.1000 F NaOH, requiring 20.00 mL to reach the end point. Calculate the percent $CaCO_3$ in the sample. In this example T is NaOH, S is HCl, and U is $CaCO_3$.

1) Write a balanced equation for the reaction of S with U and establish the stoichiometric ratio.

$$CaCO_3 + 2HCl \rightarrow CaCl_2 + CO_2(g) + H_2O$$

mol $CaCO_3$ = ½(mol HCl consumed)

2) Write a balanced equation for the reaction of T with the excess S and establish the stoichiometric ratio.

$$HCl + NaOH \rightarrow NaCl + H_2O$$

mol excess HCl = mol NaOH

3) Use the expression for the total moles of HCl added to solve for the moles of U, and finally the %U

total mol HCl = mol for $CaCO_3$ + mol in excess

total mol HCl = 2(mol $CaCO_3$) + mol NaOH

$V_{HCl}F_{HCl} = 2(\text{mol } CaCO_3) + V_{NaOH}F_{NaOH}$

mol $CaCO_3 = (V_{HCl}F_{HCl} - V_{NaOH}F_{NaOH})/2$

mol $CaCO_3$ = {(0.1000 L)(0.1000 mol/L) - (0.02000 L)(0.1000 mol/L)}/2

mol $CaCO_3$ = 0.004000 mol

mass $CaCO_3$ = (0.004000 mol)(100. 1g/mol) = 0.4004 g

%$CaCO_3$ = (0.4004 g)(100)/(1.000 g) = 40.04%

The above covers most of the types of calculations that must be performed when using titrations for quantitative analyses.

This section will be completed by asking the question - what happens if the end point is not the same as the equivalence point? Obviously an error is made. Sometimes the magnitude of that error can be calculated from theory. If the moles of titrant needed to reach the equivalence point and the moles of titrant added at the end point can both be calculated, then

$$\%\text{error} = 100\{(\text{mol T})_{e.p.} - (\text{mol T})_{eq.pt.}\}/(\text{mol T})_{eq.pt.} \quad (1\text{-}62)$$

or

$$\%\text{error} = 100\{\text{mL T}_{e.p.} - \text{mL T}_{eq.pt.}\}/\text{mL T}_{eq.pt.} \quad (1\text{-}63)$$

These equations may be useful later.

1.6 PROBLEMS

Sections 1.1 and 1.2: Statistical Treatment of Data

1-1) Compute the mean, median, range, absolute and relative standard deviations, and the 80% confidence interval for the following set of numbers: 73.8, 73.5, 74.2, 74.1, 73.6, and 73.5. Ans: 73.8, 73.7, 0.7, 0.3_1, $4._1 x10^{-3}$, and 73.6 - 74.0.

1-2) A group of students is asked to read a buret and produces the following data set: 31.45, 31.48, 31.46, 31.46, 31.44, 31.47, and 31.46 mL. Calculate the mean, percent relative standard deviation, and 95% confidence interval of this data set. Ans: 31.46 mL, 0.04_1%, 31.45 - 31.47 mL.

1-3) The following data set is available: 17.93, 17.77, 17.47, 17.82, 17.88. Calculate its mean, absolute standard deviation, and confidence interval at both the 90% and 95% level. Ans: 17.85, 0.07_0, 17.77 - 17.93, 17.77, 0.18, 17.55 - 17.99.

1-4) Calculate the mean, relative standard deviation in ppt, and the 90% confidence interval of the following data set: 41.29, 41.31, 41.30, 41.29, 41.35, 41.30, 41.28. Consult a table of Q values, if appropriate. Ans: 41.30, 0.2_5, 41.29 - 41.31.

1-5) Report the decimal value of X and its absolute standard deviation. The standard deviations shown are all absolute deviations. Ans: -11 ±2.

X=(13.50±0.08-14.50±0.06)(6.08±0.02)/(2.342±0.09-1.786±0.03)

1-6) What are the values of Z and its absolute and relative standard deviations? All standard deviations shown are absolute. Ans: 0.67, ±0.04, ±0.06.

Z=(808±7/378±5) - (6.59±0.09/4.48±0.05)

1-7) What are the values of X and its absolute standard deviation? All standard deviations shown are absolute. Ans: $73._3 \pm 3._6$ or 73±4

$X = (31.6 \pm 0.9)(1.324 \pm 0.018)^3$

1-8) Calculate the values of Y and its relative standard deviation. All standard deviations shown are absolute. Ans: 0.413, 0.015.

Y = log(100.0±0.5/38.6±0.5)

1-9) You are preparing dilutions of a stock Zn(II) solution using a buret and volumetric flasks. If the concentration of the stock Zn(II) solution is (0.01271±0.00003) M and if you deliver 15.38 mL of stock solution from the buret into a 100 mL volumetric flask, what will be the concentration of Zn(II) in the flask and its absolute standard deviation? Assume that the standard deviation of reading a buret is ±0.02 mL and of filling a 100 mL volumetric flask is ±0.03 mL and remember that the volume delivered from the buret is the result of *two* readings, each ±0.02 mL. Ans: $(1.955 \pm 0.006) \times 10^{-3}$ M.

Problems 1-17, 1-19, 1-20, and 1-26 below also involve propagation of error calculations. These should be done after the other material in the problem has been mastered.

Section 1.4: Concentration of Solutions and Dilution

1-10) Calculate the volume of glacial acetic acid (100%) needed to prepare 500.0 mL of a 0.250 F solution. The density of the reagent is 1.05 g/mL. Ans: 7.15 mL.

1-11) Concentrated reagent ammonia is 57.6% by wt. NH_3 and has a density of 0.90 g/mL. What is the formality of ammonia in a solution prepared by addition of 10.0 mL of the concentrated reagent to distilled water followed by dilution to 2.000 L? If only 3.0% of the ammonia reacts to form NH_4^+ and OH^-, what is the molarity of the ammonia in the solution? Ans: 0.15_2 F, 0.14_8 M.

1-12) Calculate the formality of H_2SO_4 in the concentrated reagent acid if the percent by weight of H_2SO_4 is 96.0% and the density is 1.81 g/mL. Ans: 17.7 mol/L.

1-13) A certain acid (GFM = 100 g/mol) is 70.0% by wt. and has a density of 1.69 g/mL. You wish to prepare 2.00 L of a 0.100 F solution of this acid. If you were preparing the solution by weight, how much reagent acid would you need to weigh out? If you were preparing it by volume, what volume of reagent acid would you need? Ans: 28.6 g, 16.9 mL.

1-14) The upper allowable limit for the concentration of Pb(II) in drinking water is 5.0 ppb. What is this in mol/L? Ans: 2.4×10^{-8}.

1-15) A 10.0 mL sample of tap water is added to a 25.0 mL volumetric flask, 10 mL of a special buffer is added, and distilled water is used to dilute to the mark. This solution is analyzed and found to contain 2.36×10^{-6} M fluoride. What is the concentration of fluoride in tap water in both mol/L and ppm? Ans: 5.90×10^{-6} M, 0.112 ppm.

1-16) What mass of KI must be weighed out to prepare 500 mL of 10.0 ppm I^- solution (you may use the simple definition of ppm)? Ans: 6.5 mg.

1-17) What will be the expected precision of the concentration calculated in Problem 16 above if the standard deviation of weighing is ±0.1 mg (remember that a single mass is the result of two weighings), and of using a 500 mL volumetric flask is ±0.05 mL? Assume that the gram formula masses are accurate. Ans: $\pm 0.2_2$ ppm.

1-18) What mass of KI would have to be weighed out to prepare the same solution as in Problem 1-16 above (500 mL of 10.0 ppm I^-), if the original salt were dissolved in 100 mL of water and 10.0 mL of that solution were diluted to 500 mL? Ans: 65.4 mg.

1-19) What would be the expected precision (abs. std. dev.) in the concentration of the solution prepared in Problem 1-18? In addition to the data given in Problem 1-17 above, assume that the precision of using a 10 mL pipet is ±0.01 mL and that of filling a 100 mL volumetric flask is ±0.02 mL. Ans: 0.02_4 ppm.

1-20) Dilute solutions (in the low ppm and the ppb range) cannot be stored for any period of time because adsorption on the walls of the storage container changes the component concentrations. Thus many stock solutions are made up at the 1000 ppm concentration level. What mass of lithium bromide would have to be weighed out to prepare 1.00 L of a 1000 ppm stock solution of Li (use the simple definition of ppm)? If one were to prepare 1.00 L of a 1.00 ppb solution of Li directly by weighing lithium bromide, what mass of the salt would have to be weighed out? If 1.00 L of a 1.00 ppb Li solution were to be prepared by diluting 1.00 mL of the 1000 ppm stock solution to 1.00 L followed by diluting 1.00 mL of this solution to 1.00 L, what would be the expected absolute precision of the final concentration? Assume one can weight to ±0.2 mg, can fill a 1.00 L volumetric flask to ±0.1 mL, and can use a 1.00 mL pipet with ±0.02 mL precision. What procedure establishes the overall precision of the final concentration? Ans: 12.51 g, 12.51 μg, ±0.03 ppb, the use of the 1.00 mL pipet.

Section 1.5: Titrations

1-21) A solution of NaOH is standardized by titrating 0.786 g of KHP. The titration requires 37.8 mL. A 0.429 g sample of a pure monoprotic acid is titrated with the same base, requiring 31.7 mL. What is the molecular mass of the monoprotic acid? Ans: 133 g/mol.

1-22) Concentrated reagent grade ammonia is reported to be 14.8 F. However, it can lose strength rapidly if not tightly stoppered and should always be checked if needed for exact work. HCl is standardized against primary standard Na_2CO_3. 0.2500 g of the standard requires 40.00 mL of HCl. 10.00 mL of the conc. NH_3 are diluted to 1.000 L. 25.00 mL of the diluted ammonia require 26.39 mL of the HCl for titration. What is the formality of the conc. reagent NH_3? Ans: 12.45 F.

1-23) The acetic acid in a vinegar sample is determined as follows: NaOH is standardized by titrating a 0.8323 g sample of KHP. The standardization requires 43.34 mL. A 2.272 g sample of vinegar is diluted to 50.00 mL and titrated with NaOH, requiring 39.87 mL. Calculate the percentage by weight of acetic acid in the vinegar sample. Ans: 9.910%.

1-24) A solution of HCl is standardized by titration against 0.277 g of primary standard Na_2CO_3. The titration requires 43.4 mL of HCl. A solid sample weighing 0.183 g and containing sodium oxide is dissolved in water ($Na_2O + H_2O \rightarrow 2NaOH$) and titrated with HCl, requiring 47.3 mL. What is the percentage of Na_2O in the sample? Ans: 96.5%.

1-25) The percentage of $CaCO_3$ in a 1.000 g sample was determined by adding 100.0 mL of 0.1000 F HCl (an excess) to dissolve the sample.

$$CaCO_3 + 2H^+ \rightarrow Ca^{2+} + H_2O + CO_2(g)$$

The solution was boiled to drive off all the CO_2 and the remaining HCl was titrated with 0.1000 F NaOH, requiring 25.00 mL. Calculate the percentage of $CaCO_3$ in the sample. Ans: 37.53%

1-26) The hydrochloric acid concentration in a solution is determined by titration with NaOH. 50.00 mL of sample require 27.31 mL of NaOH. The NaOH is standardized against oxalic acid ($H_2C_2O_4$). 0.4000 g of the acid require 38.77 mL of NaOH (the acid is diprotic). What is the formality of the HCl in the sample? If a 50 mL pipet can be used with a precision of ±0.03 mL, a balance with ±0.2 mg, and a buret with ±0.02 mL, what is the expected absolute precision of the formality? Remember that each mass is the result of two weighings and each volume from a buret is the result of two readings. Ans: 0.1252 F, ±0.0002 F (±0.000195 F).