Chapter 8:
Sampling, Standardization, and Calibration

A chemical analysis uses only a small fraction of the available sample, the process of sampling is a very important operation.

Knowing how much sample to collect and how to further subdivide the collected sample to obtain a laboratory sample is vital in the analytical process.

Statistical methods are used to aid in the selection of a representative sample.

The analytical sample must be processed in a dependable manner that maintains sample integrity without losing sample or introducing contaminants.

Many laboratories use the automated sample handling methods.

8A Analytical Samples and Methods

Types of Samples and Methods

Quantitative methods are traditionally classified as

gravimetric methods,

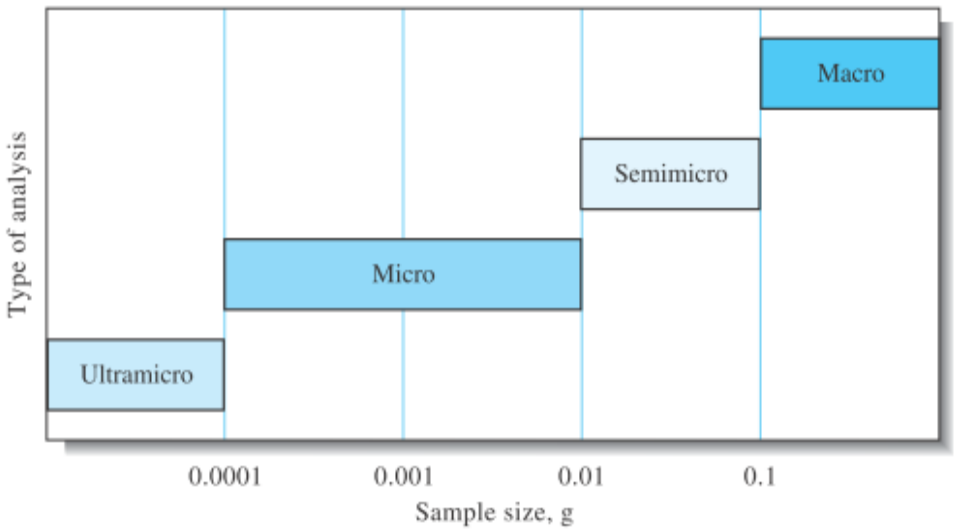
volumetric methods, and

instrumental methods.

Other methods are based on the size of the sample and the level of the constituents.

Sample Size

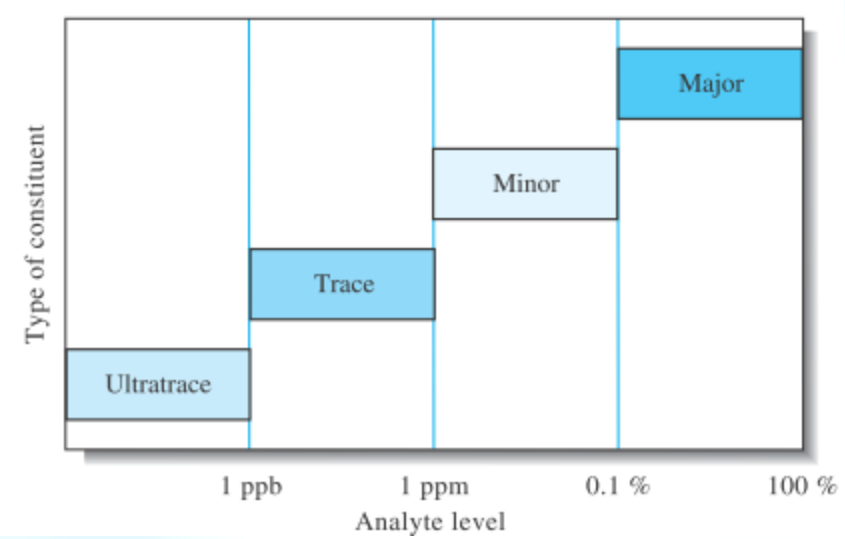
Techniques for handling very small samples are quite different from those for treating macro samples.



Sample Size	Type of Analysis
> 0.1 g	Macro
0.01 to 0.1 g	Semimicro
0.0001 to 0.01 g	Micro
$< 10^{-4}$ g	Ultramicro

Constituent Types

- In some cases, analytical methods are used to determine *major constituents*, which are those present in the range of 1 to 100% by mass.
- Species present in the range of 0.01 to 1% are usually termed *minor constituents*.
- Those present in amounts between 100 ppm (0.01%) and 1 ppb are called *trace constituents.*
- Components present in amounts lower than 1 ppb are usually considered to be *ultratrace constituents*.



Analyte Level	Type of Constituent
1 to 100%	Major
0.01 (100 ppm) to 1%	Minor
1 ppb to 100 ppm	Trace
< 1 ppb	Ultratrace

Figure 8-2 Classification of constituent types by analyte level.

➢A general problem in trace procedures is that the reliability of results usually decreases dramatically with a decrease in analyte level.

➢ The relative standard deviation between laboratories increases as the level of analyte decreases.

➢ At the ultratrace level of 1 ppb, interlaboratory error (%RSD) is nearly 50%. At lower levels, the error approaches 100%.

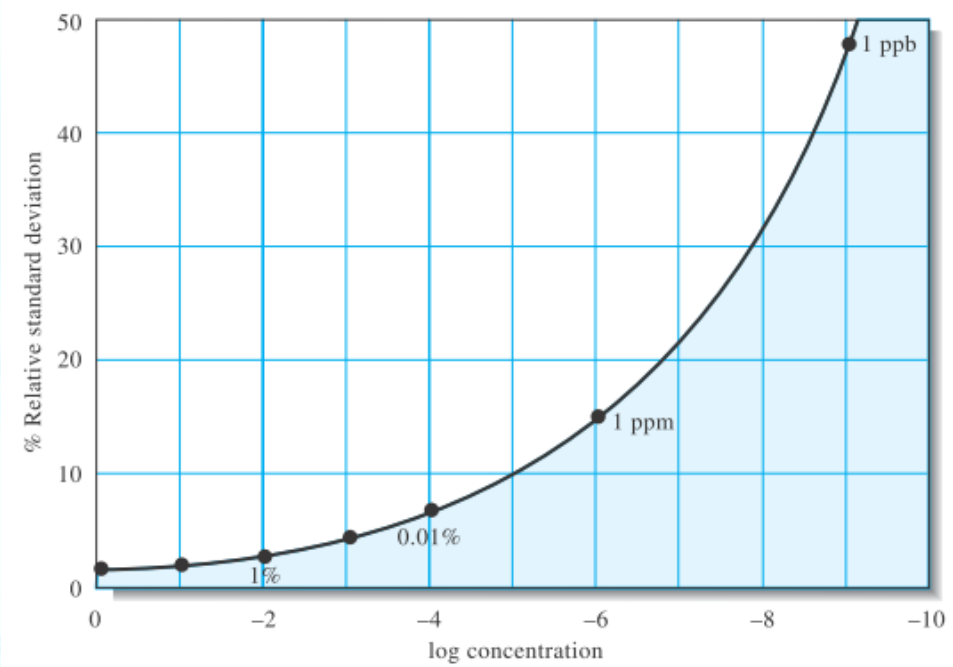


Figure 8-3 Inter-laboratory error as a function of analyte concentration.

Real Samples

- The analysis of real samples is complicated by the presence of the sample matrix.
- The matrix can contain species with chemical properties similar to the analyte.
- If the interferences are caused by extraneous species in the matrix, they are often called matrix effects.
- Such effects can be induced not only by the sample itself but also by the reagents and solvents used to prepare the samples for the determination.

Samples are analyzed, but constituents or concentrations are determined.

8B Sampling

- The process by which a representative fraction is acquired from a material of interest is termed sampling. (e.g. a few milliliters of water from a polluted lake)
- It is often the most difficult aspect of an analysis.
- Sampling for a chemical analysis necessarily requires the use of statistics because conclusions will be drawn about a much larger amount of material from the analysis of a small laboratory sample.

8B-1 Obtaining a Representative Sample

- The items chosen for analysis are often called **sampling units** or **sampling increments**.
- The collection of sampling units or increments is called the **gross sample**.
- For laboratory analysis, the gross sample is usually reduced in size and homogenized to create the **laboratory sample**.
- The composition of the gross sample and the laboratory sample must closely resemble the average composition of the total mass of material to be analyzed.

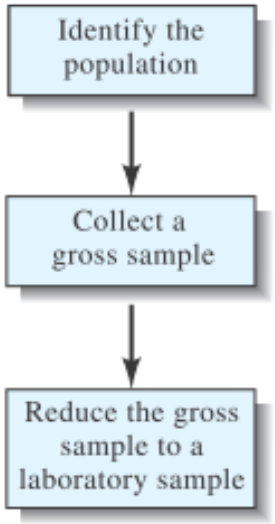


Figure 8-4 Steps in obtaining a laboratory sample. The laboratory sample consists of a few grams to at most a few hundred grams. It may constitute as little as 1 part in 10^7 -10^8 of the bulk material.

Statistically, the goals of the sampling process are:

1. To obtain a mean analyte concentration that is an unbiased estimate of the population mean. This goal can be realized only if all members of the population have an equal probability of being included in the sample.
2. To obtain a variance in the measured analyte concentration that is an unbiased estimate of the population variance so that valid confidence limits can be found for the mean, and various hypothesis tests can be applied. This goal can be reached only if every possible sample is equally likely to be drawn.

Both goals require obtaining a random sample.

A randomization procedure may be used wherein the samples are assigned a number and then a sample to be tested is selected from a table of random numbers.

For example, suppose our sample is to consist of 10 pharmaceutical tablets to be drawn from 1000 tablets off a production line. One way to ensure the sample is random is to choose the tablets to be tested from a table of random numbers. These can be conveniently generated from a random number table or from a spreadsheet as is shown in **Figure 8-5. Here, we** would assign each of the tablets a number from 1 to 1000 and use the sorted random numbers in column C of the spreadsheet to pick tablet 16, 33, 97, etc. for analysis.

	A	B	C	D	E
1	**Spreadsheet to generate random numbers between 1 and 1000**				
2		Random Numbers	Sorted Numbers		
3		97	16		
4		382	33		
5		507	97		
6		33	268		
7		511	382		
8		16	507		
9		268	511		
10		810	810		
11		934	821		
12		821	934		
13					
14	**Spreadsheet Documentation**				
15	Cell B3=RAND()*(1000-1)+1				

Figure 8-5 10 random numbers are generated from 1 to 1000 using a spreadsheet. The random number function in Excel [=RAND()] generates random numbers between 0 and 1.

8B-2 Sampling Uncertainties

- Systematic errors can be eliminated by exercising care, by calibration, and by the proper use of standards, blanks, and reference materials.
- Random errors, which are reflected in the precision of data, can generally be kept at an acceptable level by close control of the variables that influence the measurements.
- Errors due to invalid sampling are unique in the sense that they are not controllable by the use of blanks and standards or by closer control of experimental variables.
- For random and independent uncertainties, the overall standard deviation s_o for an analytical measurement is related to the standard deviation of the sampling process s_s and to the standard deviation of the method s_m by the relationship

$$s_o^2 = s_s^2 + s_m^2$$

- An analysis of variance can reveal whether the *between samples variation* (sampling plus measurement variance) is significantly greater than the *within samples variation* (measurement variance).
- When $s_m \leq s_s/3$, there is no point in trying to improve the measurement precision. This result suggests that, if the sampling uncertainty is large and cannot be improved, it is often a good idea to switch to a less precise but faster method of analysis so that more samples can be analyzed in a given length of time. Since the standard deviation of the mean is lower by a factor of $\sqrt{N}$, taking more samples can improve precision.

8B-3 The Gross Sample

Ideally, the gross sample is a miniature replica of the entire mass of material to be analyzed. It is the collection of individual sampling units. It must be representative of the whole in composition and in particle-size distribution.

Size of the Gross Sample is determined by

(1) the uncertainty that can be tolerated between the composition of the gross sample and that of the whole,

(2) the degree of heterogeneity of the whole, and

(3) the level of particle size at which heterogeneity begins.

- The number of particles, N, required in a gross sample ranges from a few particles to 10^{12} particles.
- The magnitude of this number depends on the uncertainty that can be tolerated and how heterogeneous the material is.
- The need for large numbers of particles is not necessary for homogeneous gases and liquids.
- The laws of probability govern the composition of a gross sample removed randomly from a bulk of material.

As an idealized example,

- let us presume that a pharmaceutical mixture contains just two types of particles:
  - * *type A* particles containing the active ingredient and
  - * type B particles containing only an inactive filler material.

All particles are the same size. We wish to collect a gross sample that will allow us to determine the percentage of particles containing the active ingredient in the bulk material.

--Assume that the probability of randomly drawing an A type particle is p and that of randomly drawing a B type particle is (1 - p).

-- If N particles of the mixture are taken, the most probable value for the number of A type particles is pN, while the most probable number of B type part is (1 – p)N.

-- For such a binary population, the *Bernoulli equation* can be used to calculate the standard deviation of the number of A particles drawn, σA

$$\sigma_A = \sqrt{Np(1-p)}$$

The relative standard deviation σ_r of drawing A type particles is,

$$\sigma_r = \frac{\sigma_A}{Np} = \sqrt{\frac{1-p}{Np}}$$

Thus, the number of particles needed is,

$$N = \frac{1-p}{p\sigma_r^2}$$

Thus, for example, if 80% of the particles are type A (p 5 0.8) and the desired relative standard deviation is 1% (σ_r = 0.01), the number of particles making up the gross sample should be $N=1\text{-}0.8/0.8(0.01)^2 = 2500$

➢ To determine the number of particles and thus what mass we should ensure that we have a sample with the overall average percent of active ingredient P with a sampling relative standard deviation of σ_r

$$N = p(1-p)\left(\frac{d_A d_B}{d^2}\right)^2\left(\frac{P_A - P_B}{\sigma_r P}\right)^2$$

➢ The degree of heterogeneity as measured by P_A - P_B has a large influence on the number of particles required since N increases with the square of the difference in composition of the two components of the mixture. Rearranging the equation to calculate the relative standard deviation of sampling, σ_r we get

$$\sigma_r = \frac{|P_A - P_B|}{P} \times \frac{d_A d_B}{d^2}\sqrt{\frac{p(1-p)}{N}}$$

If we make the assumption that the sample mass m is proportional to the number of particles and the other quantities are constant, the product of m and σ_r should be a constant. This constant K_s is called the *Ingamells sampling constant.*

$$K_s = m \times (\sigma_r \times 100)^2$$

where the term $\sigma_r \times 100\%$ is the percent relative standard deviation.

- To simplify the problem of defining the mass of a gross sample of a multi-component mixture, assume that the sample is a hypothetical two-component mixture.
- The problem of variable particle size can be handled by calculating the number of particles that would be needed if the sample consisted of particles of a single size.
- The gross sample mass is then determined by taking into account the particle-size distribution.
- One approach is to calculate the necessary mass by assuming that all particles are the size of the largest.
- This procedure is not very efficient because it usually calls for removal of a larger mass of material than necessary.
- The mass of the sample increases directly as the volume (or as the cube of the particle diameter) so that reduction in the particle size of a given material has a large effect on the mass required for the gross sample.

EXAMPLE 8-1

A column-packing material for chromatography consists of a mixture of two types of particles. Assume that the average particle in the batch being sampled is approximately spherical with a radius of about 0.5 mm. Roughly 20% of the particles appear to be pink in color and are known to have about 30% by mass of a polymeric stationary phase attached (analyte). The pink particles have a density of 0.48 g/cm^3. The remaining particles have a density of about 0.24 g/cm^3 and contain little or no polymeric stationary phase. What mass of the material should the gross sample contain if the sampling uncertainty is to be kept below 0.5% relative?

Solution

We first compute values for the average density and percent polymer:

$$d = 0.20 \times 0.48 + 0.80 \times 0.24 = 0.288 \text{ g/cm}^3$$

$$P = \frac{(0.20 \times 0.48 \times 0.30) \text{ g polymer/cm}^3}{0.288 \text{ g sample/cm}^3} \times 100\% = 0.10\%$$

Then, substituting into Equation 8-5 gives

$$N = 0.20(1 - 0.20)\left[\frac{0.48 \times 0.24}{(0.288)^2}\right]^2\left(\frac{30 - 0}{0.005 \times 10.0}\right)^2$$

$$= 1.11 \times 10^5 \text{ particles required}$$

$$\text{mass of sample} = 1.11 \times 10^5 \cancel{\text{particles}} \times \frac{4}{3}\pi(0.05)^3 \frac{\cancel{\text{cm}^3}}{\cancel{\text{particle}}} \times \frac{0.288 \text{ g}}{\cancel{\text{cm}^3}}$$

$$= 16.7 \text{ g}$$

Sampling Homogeneous Solutions of Liquids and Gases

➢ Well-mixed solutions of liquids and gases require only a very small sample because they are homogeneous down to the molecular level. Gases can be sampled by several methods.

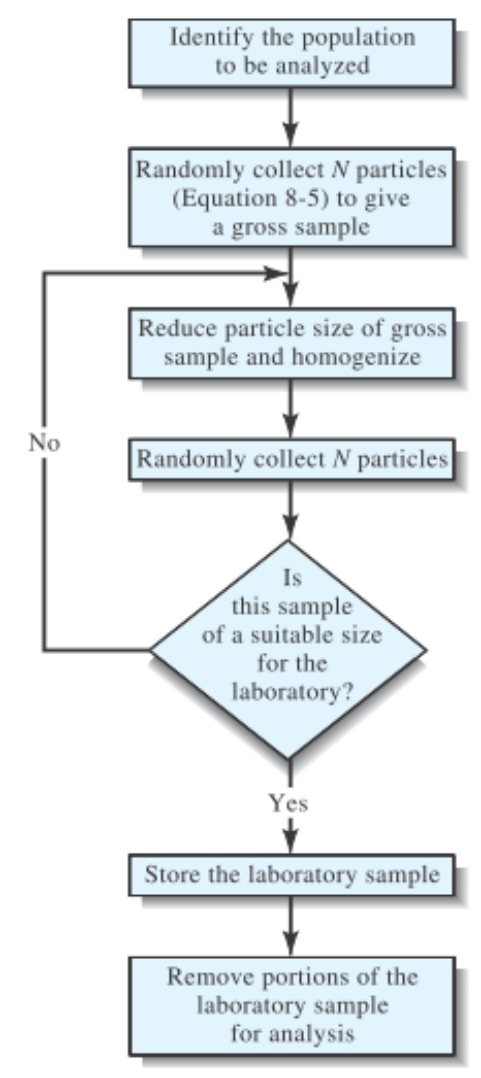
Ex., a sampling bag is simply opened and filled with the gas or gases can be trapped in a liquid or adsorbed onto the surface of a solid.

Sampling Metals and Alloys

➢ Samples of metals and alloys are obtained by sawing, milling, or drilling.

It is not safe to assume that chips of the metal removed from the surface are representative of the entire bulk.
Solid from the interior must be sampled as well.
With some materials, a representative sample can be obtained by sawing across the piece at random intervals.

Sampling Particulate Solids



- It is often difficult to obtain a random sample from a bulky particulate material.
- Random sampling can best be accomplished while the material is being transferred.
- Mechanical devices have been developed for handling many types of particulate matter.

Figure 8-6 Sampling Particulate Solids

8B-4 Preparing a Laboratory Sample

- For heterogeneous solids, the mass of the gross sample may range from hundreds of grams to kilograms or more.
- Reduction of the gross sample to a finely ground and homogeneous laboratory sample, of at most a few hundred grams, is necessary.
- this process involves a cycle of operations that includes crushing and grinding, sieving, mixing, and dividing the sample (often into halves) to reduce its mass.

EXAMPLE 8-2

A carload of lead ore containing galena (≈ 70% Pb) and other particles with little or no lead is to be sampled. From the densities (galena = 7.6 g/cm^3, other particles = 3.5 g/cm^3, average density = 3.7 g/cm^3) and rough percentage of lead, Equation 8-5 indicates that 8.45×10^5 particles are required to keep the sampling error below 0.5% relative. The particles appear spherical with a radius of 5 mm. A calculation of the sample mass required, similar to that in Example 8-1, shows that the gross sample mass should be about 1.6×10^6 g (1.8 ton). The gross sample needs to be reduced to a laboratory sample of about 100 g. How can this be done?

Solution

The laboratory sample should contain the same number of particles as the gross sample, or 8.45×10^5. The average mass of each particle, m_{avg}, is then

$$m_{avg} = \frac{100\ \text{g}}{8.45 \times 10^5\ \text{particles}} = 1.18 \times 10^{-4}\ \text{g/particle}$$

The average mass of a particle is related to its radius in cm by the equation

$$m_{avg} = \frac{4}{3}\pi r^3 \times \frac{3.7\ \text{g}}{\text{cm}^3}$$

Since $m_{avg} = 1.18 \times 10^{-4}$ g / particle, we can solve for the average particle radius r:

$$r = \left(1.18 \times 10^{-4}\ \text{g} \times \frac{3}{4\pi} \times \frac{\text{cm}^3}{3.7\ \text{g}}\right)^{1/3} = 1.97 \times 10^{-2}\ \text{cm or } 0.2\ \text{mm}$$

Thus, the sample should be repeatedly ground, mixed, and divided until the particles are about 0.2 mm in diameter.

Number of Laboratory Samples

- The number, of samples, depends on the required confidence interval and the desired relative standard deviation of the method.
- If the sampling standard deviation σ_s is known, we can use values of z from tables, to get:

$$CIfor\mu = \bar{x} \pm \frac{z\sigma_s}{\sqrt{N}}$$

- Usually, an estimate of σ_s is used with t instead of z $CIfor\mu = \bar{x} \pm \frac{ts_s}{\sqrt{N}}$

- If we divide this term by the mean value x, we can calculate the relative uncertainty σ_r that is tolerable at a given confidence level:

$$\sigma_r = \frac{ts_s}{\bar{x}\sqrt{N}}$$

If we solve Equation 8-8 for the number of samples *N, we obtain*

$$N = \frac{t^2 s_s^2}{\sigma_r^2 x^2}$$

EXAMPLE 8-3

The determination of copper in a seawater sample gave a mean value of 77.81 μg/L and a standard deviation s_s of 1.74 μg/L. (Note: the insignificant figures were retained here because these results are used below in another calculation.) How many samples must be analyzed to obtain a relative standard deviation of 1.7% in the results at the 95% confidence level?

Solution

We begin by assuming that we have an infinite number of samples, which corresponds to a t value of 1.96 at the 95% confidence level. Since $\sigma_r = 0.017$, $s_s = 1.74$, and $\bar{x} = 77.81$, Equation 8-9 gives

$$N = \frac{(1.96)^2 \times (1.74)^2}{(0.017)^2 \times (77.81)^2} = 6.65$$

We round this result to 7 samples and find the value of t for 6 degrees of freedom is 2.45. Using this t value, we then calculate a second value for N which is 10.38. Now if we use 9 degrees of freedom and $t = 2.26$, the next value is $N = 8.84$. The iterations converge with an N value of approximately 9. Note that it would be good strategy to reduce the sampling uncertainty so that fewer samples would be needed.

8 C Automated sample handling

Automated sample handling can lead to higher throughput (more analyses per unit time), higher reliability, and lower costs than manual sample handling.

Discrete (Batch) Methods

These often mimic the operations that would be performed manually.

Some discrete sample processors automate only the measurement step of the procedure or a few chemical steps and the measurement step.

Continuous Flow Methods

The sample is inserted into a flowing stream where a number of operations can be performed prior to transporting it to a flow-through detector.

These methods can perform not only sample processing operations but also the final measurement step.

Two types of continuous flow analyzers are
* the segmented flow analyzer and
* the flow injection analyzer.

•The segmented flow analyzer divides the sample into discrete segments separated by gas bubbles.

• the gas bubbles provide barriers to prevent the sample from spreading out along the tube due to dispersion processes.

• Dispersion is a band-spreading or mixing phenomenon that results from the coupling of fluid flow with molecular diffusion.

• Diffusion is mass transport due to a concentration gradient.

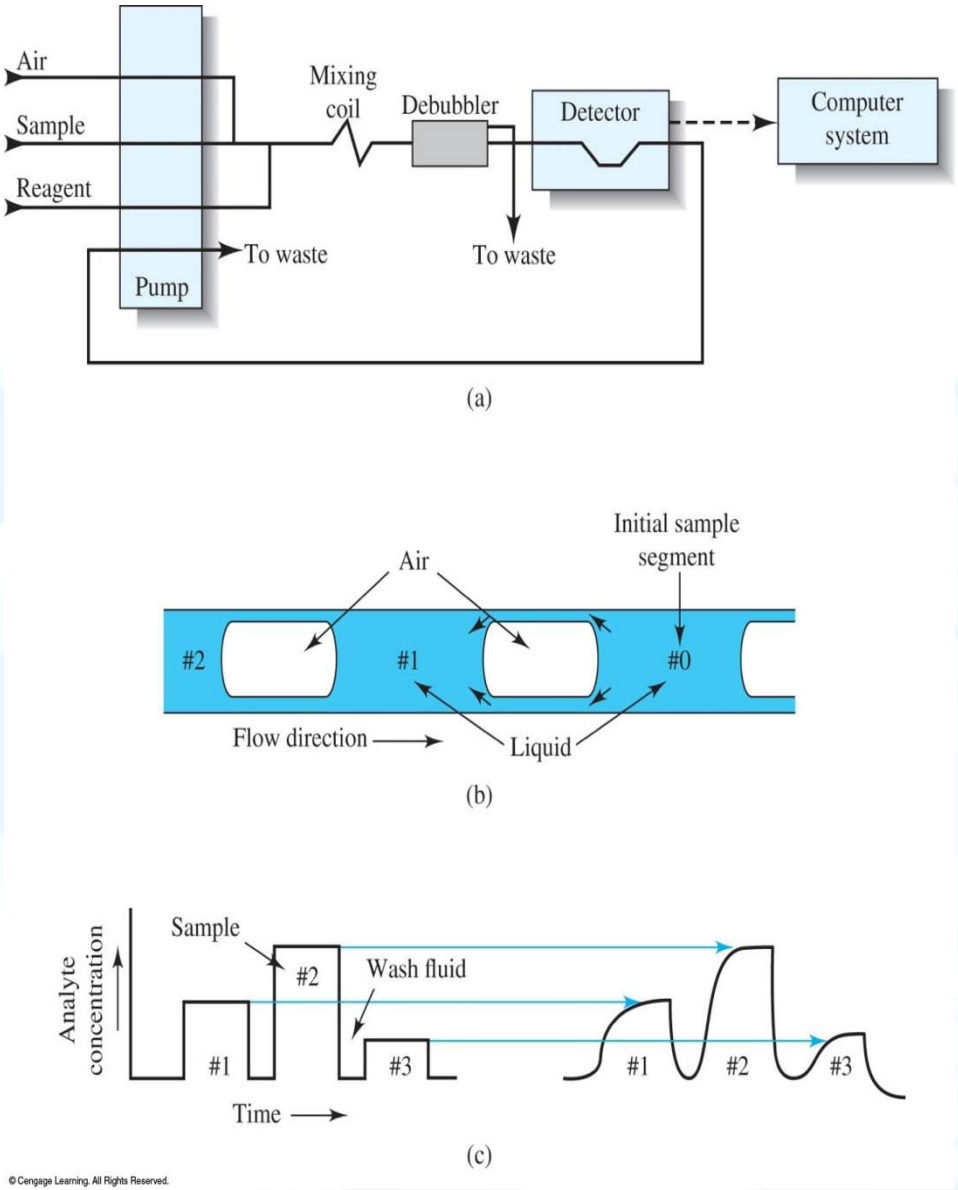


Figure 8-7 Segmented continuous flow analyzer. The segmented sample is shown in more detail in (b).The analyte concentration profiles at the sampler and at the detector are shown in (c). Normally the height of a sample peak is related to the concentration of the analyte.

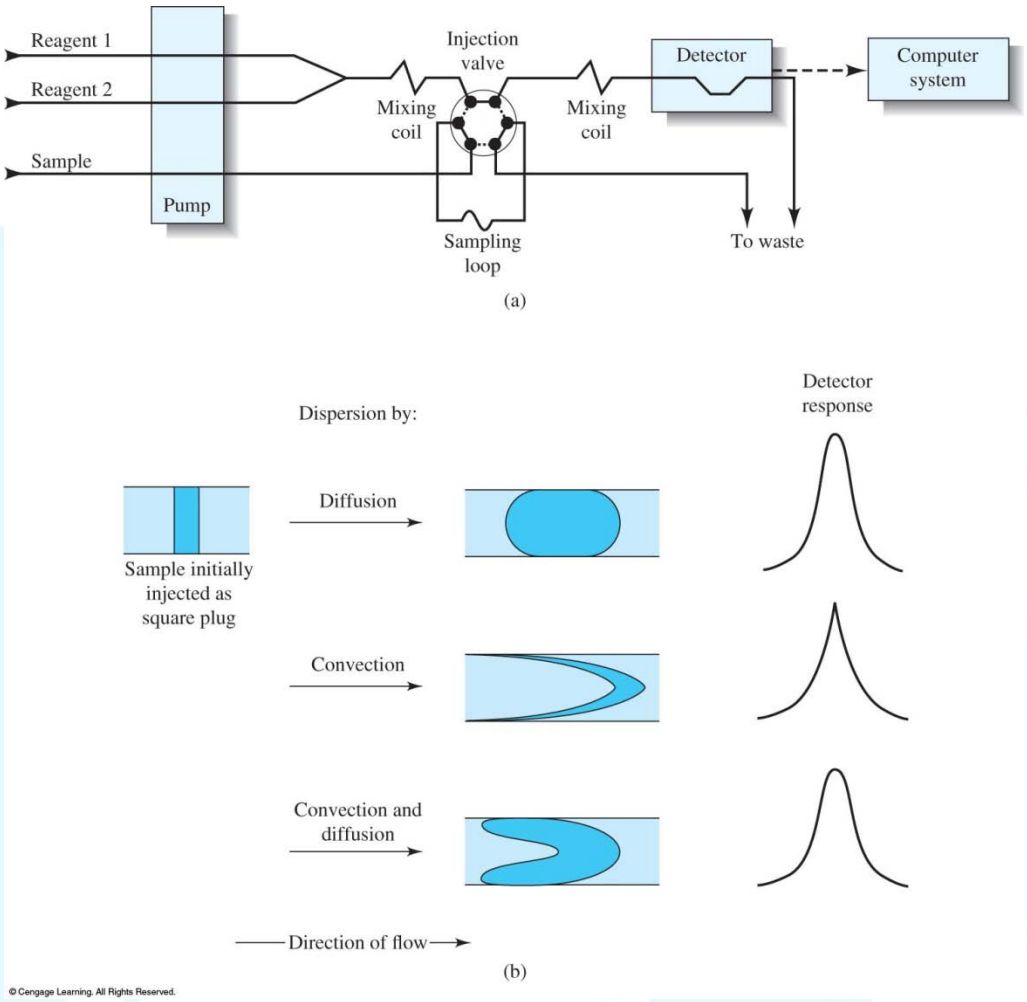


Figure 8-8 Flow injection analyzer. Samples can be processed with FIA at rates varying from 60 to 300 samples per hour. The valve, shown in the load position, also has a second inject position shown by the dotted lines. When switched to the inject position, the stream containing the reagent flows through the sample loop.

Sample and reagent are allowed to mix and react in the mixing coil before reaching the detector. In this case, the sample plug is allowed to disperse prior to reaching the detector (b). The resulting concentration profile (detector response) depends on the degree of dispersion.

8D Standardization and calibration

- Calibration determines the relationship between the analytical response and the analyte concentration, which is usually determined by the use of ***chemical standards*** prepared from purified reagents.

- To reduce interferences from other constituents in the sample matrix, called concomitants, standards are added to the analyte solution (internal standard methods or standard addition methods) or matrix matching or modification is done.

- Almost all analytical methods require calibration with chemical standards.

- Gravimetric methods and some coulometric methods are absolute methods that do not rely on calibration with chemical standards.

8D-1 Comparison with Standards

Two types of comparison methods are:

- direct comparison techniques
- titration procedures.

Direct Comparison

- Some analytical procedures involve comparing a property of the analyte with standards such that the property being tested matches or nearly matches that of the standard. This is called null comparison or isomation methods.

-Some modern instruments use a variation of this procedure to determine if an analyte concentration exceeds or is less than some threshold level. Such a comparator can be used to determine whether the threshold has been exceeded.

-e.g. A comparator to determine whether aflatoxin levels in a sample exceeds the threshold level that would indicate a toxic situation.

Titrations:

- Titrations are one of the most accurate of all analytical procedures.

- In a titration, the analyte reacts with a standardized reagent (the titrant) in a known stoichiometric manner.

- The amount of titrant is varied until chemical equivalence is reached as indicated by the color change of a chemical indicator or by the change in an instrument response. This is called the end point.

- The amount of the standardized reagent needed to achieve chemical equivalence can then be related to the amount of analyte present by means of the stoichiometry.

- Titration is thus a type of chemical comparison.

8D-2- External Standard Calibration

- A series of standard solutions is prepared separately from the sample.

- The standards are used to establish the instrument calibration function, which is obtained from analysis of the instrument response as a function of the known analyte concentration.

- The calibration function can be obtained graphically or in mathematical form.

- Generally, a plot of instrument response versus known analyte concentrations is used to produce a calibration curve, sometimes called a working curve.

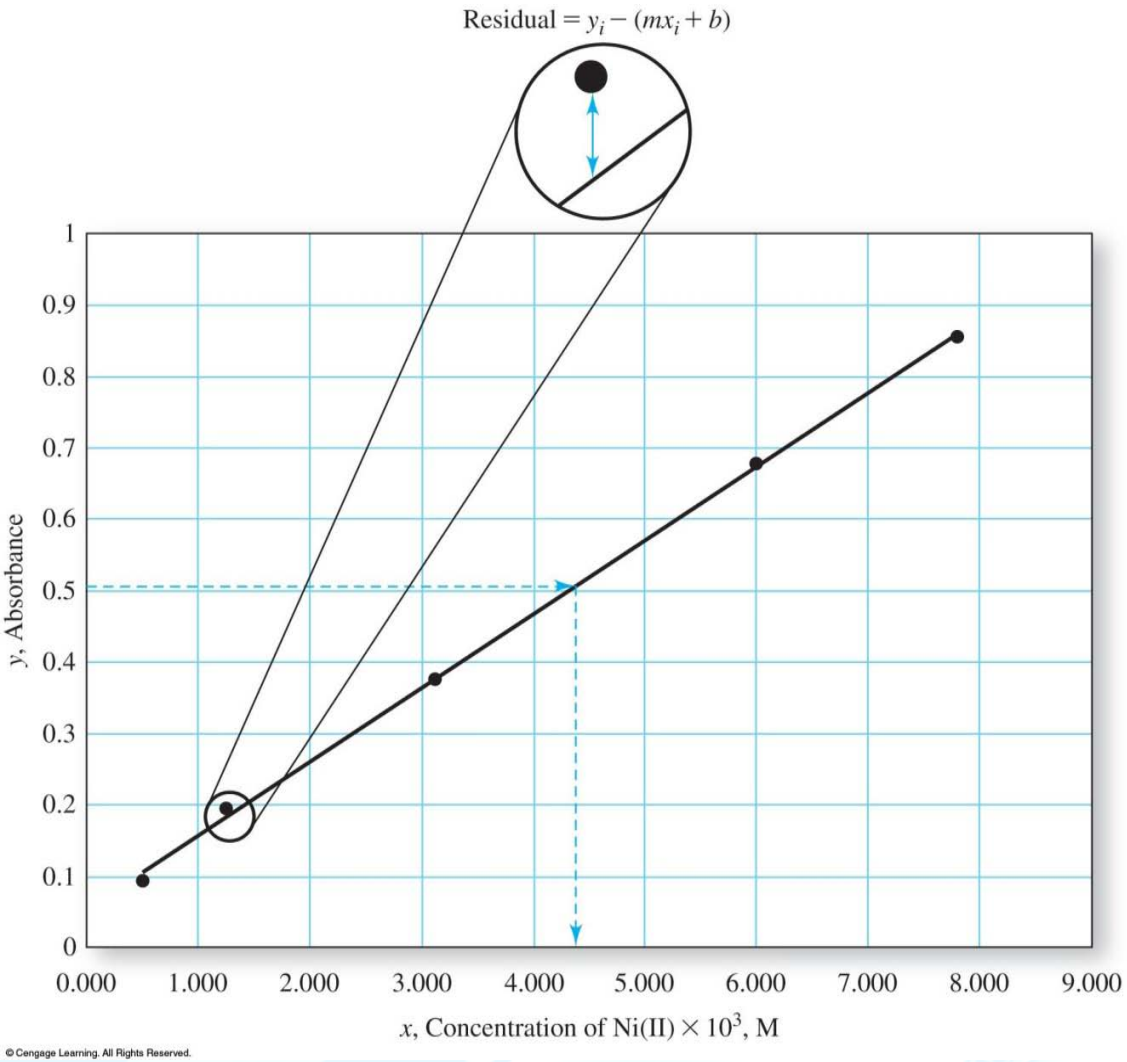


Figure 8-9
Calibration curve of absorbance versus analyte concentration for a series of standards.

-The calibration curve is used in an inverse fashion to obtain the concentration of an unknown with an absorbance of 0.505.

- The absorbance is located on the line, and then the concentration corresponding to that absorbance is obtained by extrapolating to the x-axis.

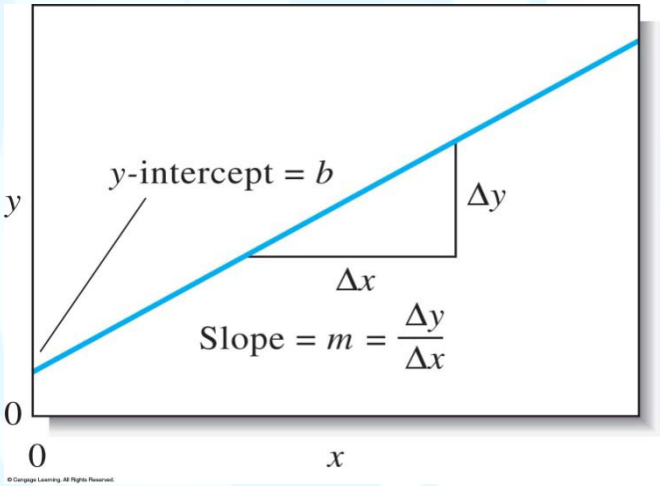
External Standard Calibration

The Least-Squares Method

Statistical methods, such as the method of least squares, are routinely used to find the mathematical equation describing the calibration function.

Two assumptions are made:

1.There is actually a ***linear relationship*** between the measured response y (absorbance) and the standard analyte concentration x.



The mathematical relationship that describes this assumption is called the regression model, which may be represented as

$$y = mx + b$$

where,

b is the y intercept (the value of y when x is zero), and m is the slope of the line.

Figure 8-10 The slope-intercept form of a straight line.

2. We also assume that any deviation of the individual points from the straight line arises from error in the measurement. That is, we assume there is no error in x values of the points (concentrations).

Whenever there is significant uncertainty in the x data, basic linear least-squares analysis may not give the best straight line in which case, a more complex correlation analysis may be used.

It may be necessary to apply different weighting factors to the points and perform a weighted least-squares analysis.

Finding the least-Squares line

The least-squares method finds the sum of the squares of the residuals SS_{resid} and minimizes the sum using calculus.

$$SS_{resid} = \sum_{i=1}^{N} \left[y_i \right| - (b + mx_i) \Big]^2$$

The slope and the intercept are defined as:

$$S_{xx} = \sum (x_i - \overline{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{N}$$

$$S_{yy} = \sum (y_i - \overline{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{N}$$

$$S_{xy} = \sum (x_i - \overline{x})(y_i - \overline{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{N}$$

where x_i and y_i are individual pairs of data for x and y, N is the number of pairs, and x and y are the average values for x and y.

From these values, one can derive the

(1) Slope of the line,

$$m = S_{xy} / S_{xx}$$

(2) Intercept,

$$b = \overline{y} - m\overline{x}$$

(3) Standard deviation about regression

$$s_r = \sqrt{\frac{S_{yy} - m^2 S_{xx}}{N-2}}$$

(4) Standard deviation of the slope

$$s_m = \sqrt{\frac{s_r^2}{S_{xx}}}$$

(5) Standard deviation of the intercept

$$s_b = s_r \sqrt{\frac{\sum x_i^2}{N\sum x_i^2 - (\sum x_i)^2}}$$

$$= s_r \sqrt{\frac{1}{N - (\sum x_i)^2 / (\sum x_i^2)}}$$

(6) Standard deviation for results obtained from the calibration curve

$$s_c = \frac{s_r}{m}\sqrt{\frac{1}{M} + \frac{1}{N} + \frac{(\overline{y_c} - \overline{y})^2}{m^2 S_{xx}}}$$

The *standard deviation about regression*, also called the *standard error of the estimate* or just the standard error, is a rough measure of the magnitude of a typical deviation from the regression line.

$$s_r = \sqrt{\frac{\sum_{i=1}^{N}[y_i - (b + mx_i)]^2}{N-2}} = \sqrt{\frac{SS_{resid}}{N-2}}$$

EXAMPLE 8-4

Carry out a least-squares analysis of the calibration data for the determination of isooctane in a hydrocarbon mixture provided in the first two columns of **Table 8-1**.

TABLE 8-1

Calibration Data for the Chromatographic Determination of Isooctane in a Hydrocarbon Mixture

Mole Percent Isooctane, x_i	Peak Area y_i	x_i^2	y_i^2	$x_i y_i$
0.352	1.09	0.12390	1.1881	0.38368
0.803	1.78	0.64481	3.1684	1.42934
1.08	2.60	1.16640	6.7600	2.80800
1.38	3.03	1.90440	9.1809	4.18140
1.75	4.01	3.06250	16.0801	7.01750
5.365	12.51	6.90201	36.3775	15.81992

Columns 3, 4, and 5 of the table contain computed values for x_i^2, y_i^2, and $x_i y_i$, with their sums appearing as the last entry in each column. Note that the number of digits carried in the computed values should be the *maximum allowed by the calculator or computer*, that is, *rounding should not be performed until the calculation is complete.*

Solution

We now substitute into Equations 8-10, 8-11, and 8-12 and obtain

$$S_{xx} = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{N} = 6.9021 - \frac{(5.365)^2}{5} = 1.14537$$

$$S_{yy} = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{N} = 36.3775 - \frac{(12.51)^2}{5} = 5.07748$$

$$S_{xy} = \Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{N} = 15.81992 - \frac{5.365 \times 12.51}{5} = 2.39669$$

Substitution of these quantities into Equations 8-13 and 8-14 yields

$$m = \frac{2.39669}{1.14537} = 2.0925 \approx 2.09$$

$$b = \frac{12.51}{5} - 2.0925 \times \frac{5.365}{5} = 0.2567 \approx 0.26$$

Thus, the equation for the least-squares line is

$$y = 2.09x + 0.26$$

Substitution into Equation 8-15 yields the standard deviation about regression,

$$s_r = \sqrt{\frac{S_{yy} - m^2 S_{xx}}{N - 2}} = \sqrt{\frac{5.07748 - (2.0925)^2 \times 1.14537}{5 - 2}} = 0.1442 \approx 0.14$$

and substitution into Equation 8-16 gives the standard deviation of the slope,

$$s_m = \sqrt{\frac{s_r^2}{S_{xx}}} = \sqrt{\frac{(0.1442)^2}{1.14537}} = 0.13$$

Finally, we find the standard deviation of the intercept from Equation 8-17:

$$s_b = 0.1442\sqrt{\frac{1}{5 - (5.365)^2/6.9021}} = 0.16$$

EXAMPLE 8-5

The calibration curve found in Example 8-4 was used for the chromatographic determination of isooctane in a hydrocarbon mixture. A peak area of 2.65 was obtained. Calculate the mole percent of isooctane in the mixture and the standard deviation if the area was (a) the result of a single measurement and (b) the mean of four measurements.

Solution

In either case, the unknown concentration is found from rearranging the least-squares equation for the line, which gives

$$x = \frac{y - b}{m} = \frac{y - 0.2567}{2.0925} = \frac{2.65 - 0.2567}{2.0925} = 1.144 \text{ mol \%}$$

(a) Substituting into Equation 8-18, we obtain

$$s_c = \frac{0.1442}{2.0925}\sqrt{\frac{1}{1} + \frac{1}{5} + \frac{(2.65 - 12.51/5)^2}{(2.0925)^2 \times 1.145}} = 0.076 \text{ mole \%}$$

(b) For the mean of four measurements,

$$s_c = \frac{0.1442}{2.0925}\sqrt{\frac{1}{4} + \frac{1}{5} + \frac{(2.65 - 12.51/5)^2}{(2.0925)^2 \times 1.145}} = 0.046 \text{ mole \%}$$

Interpretation of least-Squares results

The *sum of the squares of the residuals*, SS_{resid}, measures the variation in the observed values of the dependent variable (y values) that are not explained by the presumed linear relationship between x and y.

$$SS_{resid} = \sum_{i=1}^{N} [y_i - (b + mx_i)]^2$$

$$SS_{tot} = S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{N}$$

The coefficient of determination (R^2) measures the fraction of the observed variation in y that is explained by the linear relationship.

$$R^2 = 1 - \frac{SS_{resid}}{SS_{tot}}$$

The difference between SS_{tot} and SS_{resid} is the sum of the squares due to regression, SS_{regr}.

$$SS_{regr} = SS_{tot} - SS_{resid}$$

$$R^2 = \frac{SS_{regr}}{SS_{tot}}$$

* A significant regression is one in which the variation in the y values due to the presumed linear relationship is large compared to that due to error (residuals).

* The F value gives us an indication of the significance of the regression. When the regression is significant, a large value of F occurs.

EXAMPLE 8-6

Find the coefficient of determination for the chromatographic data of Example 8-4.

Solution

For each value of x_i, we can find a predicted value of y_i from the linear relationship. Let us call the predicted values of y_i, $\hat{y}_i$. We can write $\hat{y}_i = b + mx_i$ and make a table of the observed y_i values, the predicted values $\hat{y}_i$, the residuals $y_i - \hat{y}_i$, and the squares of the residuals $(y_i - \hat{y}_i)^2$. By summing the latter values, we obtain SS_{resid} as shown in Table 8-2.

TABLE 8-2

Finding the Sum of the Squares of the Residuals

	x_i	y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
	0.352	1.09	0.99326	0.09674	0.00936
	0.803	1.78	1.93698	−0.15698	0.02464
	1.08	2.60	2.51660	0.08340	0.00696
	1.38	3.03	3.14435	−0.11435	0.01308
	1.75	4.01	3.91857	0.09143	0.00836
Sums	5.365	12.51			0.06240

From Example 8-4, the value of $S_{yy} = 5.07748$. Hence,

$$R^2 = 1 - \frac{SS_{resid}}{SS_{tot}} = 1 - \frac{0.0624}{5.07748} = 0.9877$$

This calculation shows that over 98% of the variation in peak area can be explained by the linear model.

We can also calculate SS_{regr} as

$$SS_{regr} = SS_{tot} - SS_{resid} = 5.07748 - 0.06240 = 5.01508$$

Let us now calculate the F value. There were five xy pairs used for the analysis. The total sum of the squares has 4 degrees of freedom associated with it since one is lost in calculating the mean of the y values. The sum of the squares due to the residuals has 3 degrees of freedom because two parameters m and b are estimated. Hence SS_{regr} has only 1 degree of freedom since it is the difference between SS_{tot} and SS_{resid}. In our case, we can find F from

$$F = \frac{SS_{regr}/1}{SS_{resid}/3} = \frac{5.01508/1}{0.0624/3} = 241.11$$

This very large value of F has a very small chance of occurring by random chance, and therefore, we conclude that this is a significant regression.

Transformed Variables

Linear least squares gives best estimates of the transformed variables, but these may not be optimal when transformed back to obtain estimates of the original parameters.

For the original parameters, nonlinear regression methods may give better estimates.

TABLE 8-3

Transformations to Linearize Functions

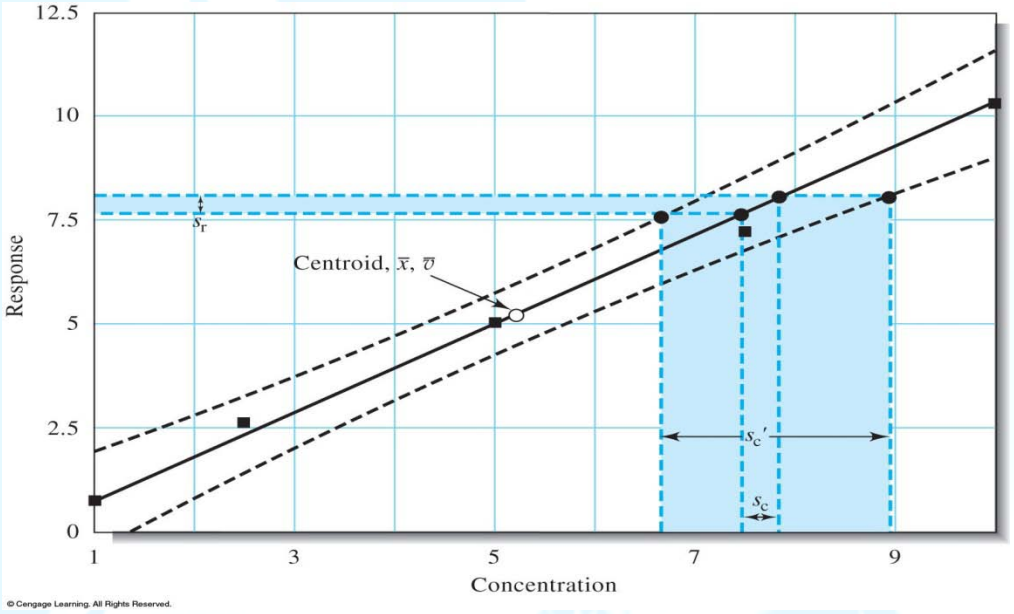
Function	Transformation to Linearize	Resulting Equation
Exponential: $y = be^{mx}$	$y' = \ln(y)$	$y' = \ln(b) + mx$
Power: $y = bx^m$	$y' = \log(y), x' = \log(x)$	$y' = \log(b) + mx'$
Reciprocal $y = b + m\left(\frac{1}{x}\right)$	$x' = \frac{1}{x}$	$y = b + mx'$

Errors in External Standard Calibration

- When external standards are used, it is assumed that, when the same analyte concentration is present in the sample and in the standard, the same response will be obtained.
- The raw response from the instrument is usually not used.
- Instead, the raw analytical response is corrected by measuring a blank. The ideal blank is identical to the sample but without the analyte.
- A real blank is either a solvent blank, containing the same solvent in which the sample is dissolved, or a reagent blank, containing the solvent plus all the reagents used in sample preparation.
- Systematic errors can also occur during the calibration process.
- To avoid systematic errors in calibration, standards must be accurately prepared, and their chemical state must be identical to that of the analyte in the sample.
- The standards should be stable in concentration, at least during the calibration process.
- *Random errors can also influence the accuracy of results obtained from calibration curves.*

Figure 8-11 Shown here is a calibration curve with confidence limits.

Measurements made near the center of the curve will give less uncertainty in analyte concentration than those made at the extremes.



Minimizing Errors in Analytical Procedures

The overall accuracy and precision of an analysis is not limited to the measurement step and might instead be limited by factors such as sampling, sample preparation, and calibration.

Separations

Sample cleanup by separation methods is an important way to minimize errors from possible interferences in the sample matrix.

Techniques such as filtration, precipitation, dialysis, solvent extraction, volatilization, ion exchange, and chromatography can be used.

In most cases, separations may be the only way to eliminate an interfering specimen.

Saturation, Matrix Modification, and Masking

*The saturation method involves adding the interfering species to all the samples, standards, and blanks so that the interference effect becomes independent of the original concentration of the interfering species in the sample.

*A matrix modifier is a species, not itself an interfering species, added to samples, standards, and blanks in sufficient amounts to make the analytical response independent of the concentration of the interfering species.

*Sometimes, a masking agent is added that reacts selectively with the interfering species to form a complex that does not interfere.

Dilution and Matrix Matching

*The dilution method can sometimes be used if the interfering species produces no significant effect below a certain concentration level.

*The matrix-matching method attempts to duplicate the sample matrix by adding the major matrix constituents to the standard and blank solutions.

*Errors in procedures can be minimized by saturating with interfering species, by adding matrix modifiers or masking agents, by diluting the sample, or by matching the matrix of the sample.

Internal Standard Methods

*An internal standard is a reference species, chemically and physically similar to the analyte, that is added to samples, standards, and blanks.

*The ratio of the response of the analyte to that of the internal standard is plotted versus the concentration of analyte.

*In the internal standard method, a known amount of a reference species is added to all the samples, standards, and blanks.

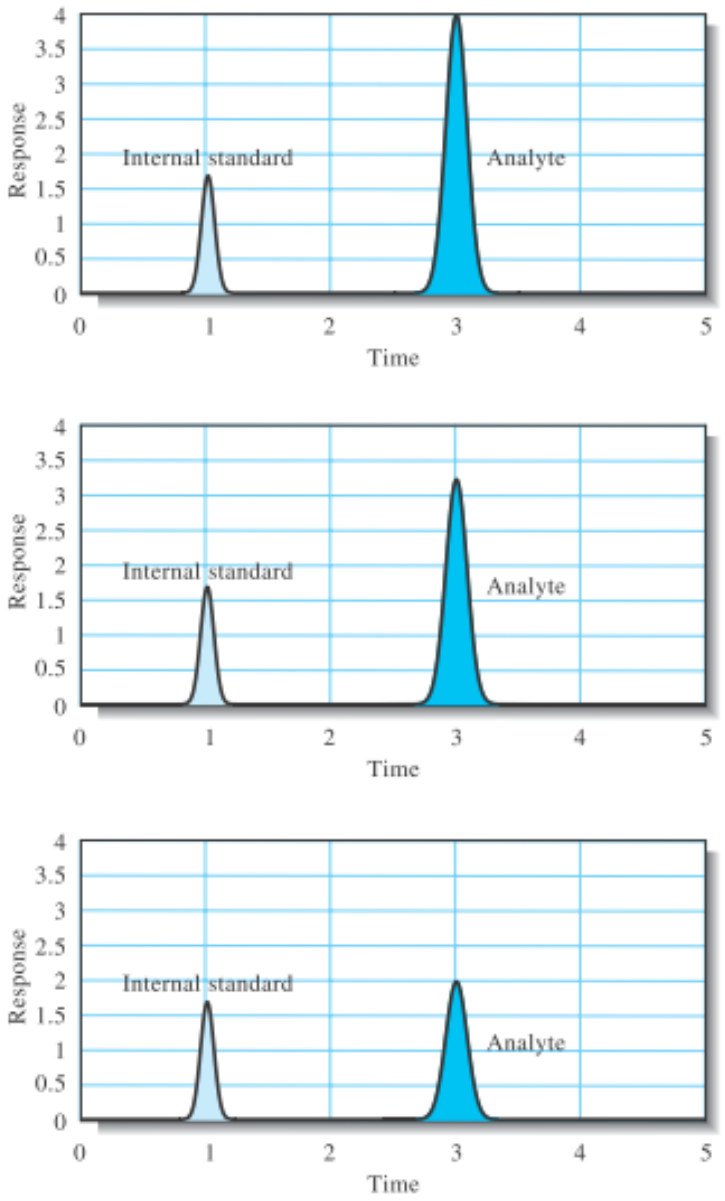
*The response signal is then not the analyte signal itself but the ratio of the analyte signal to the reference species signal.

*A calibration curve is prepared where the y-axis is the ratio of responses and the x-axis is the analyte concentration in the standards as usual.

*This method can compensate for certain types of errors if these influence both the analyte and the reference species to the same proportional extent.

*The calibration curve plots the ratio of the analyte signal to the internal standard signal against the concentration of the analyte.

Figure 8-12 Illustration of the internal standard method.



EXAMPLE 8-7

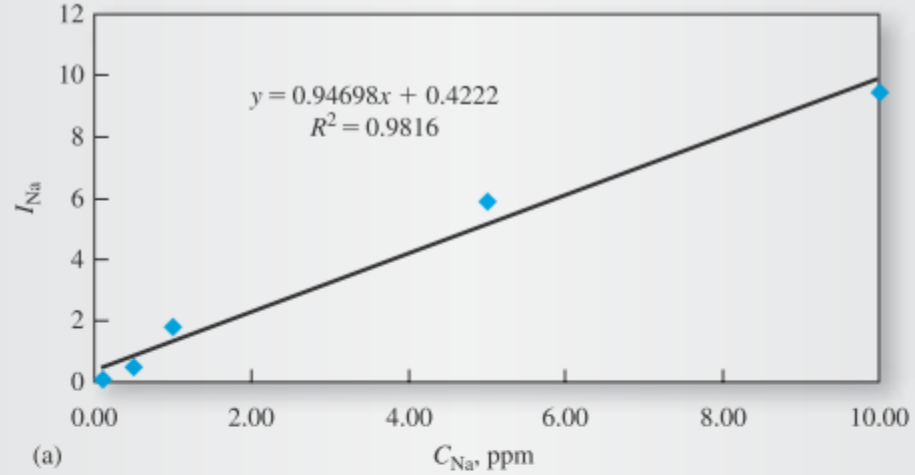
The intensities of flame emission lines can be influenced by a variety of instrumental factors, including flame temperature, flow rate of solution, and nebulizer efficiency. We can compensate for variations in these factors by using the internal standard method. Thus, we add the same amount of internal standard to mixtures containing known amounts of the analyte and to the samples of unknown analyte concentration. We then take the ratio of the intensity of the analyte line to that of the internal standard. The internal standard should be absent in the sample to be analyzed.

In the flame emission determination of sodium, lithium is often added as an internal standard. The following emission intensity data were obtained for solutions containing Na and 1000 ppm Li.

(*continued*)

c_{Na}, ppm	Na intensity, I_{Na}	Li intensity, I_{Li}	I_{Na}/I_{Li}
0.10	0.11	86	0.00128
0.50	0.52	80	0.0065
1.00	1.8	128	0.0141
5.00	5.9	91	0.0648
10.00	9.5	73	0.1301
Unknown	4.4	95	0.0463

A plot of the Na emission intensity versus the Na concentration is shown in **Figure 8-13a**. Note that there is some scatter in the data and the R^2 value is 0.9816. In **Figure 8-13b**, the ratio of the Na to Li emission intensities is against the Na concentration. Note that the linearity is improved as indicated by the R^2 value of 0.9999. The unknown intensity ratio (0.0463) is then located on the curve, and the concentration of Na corresponding to this ratio is found to be 3.55 ± 0.05 ppm.



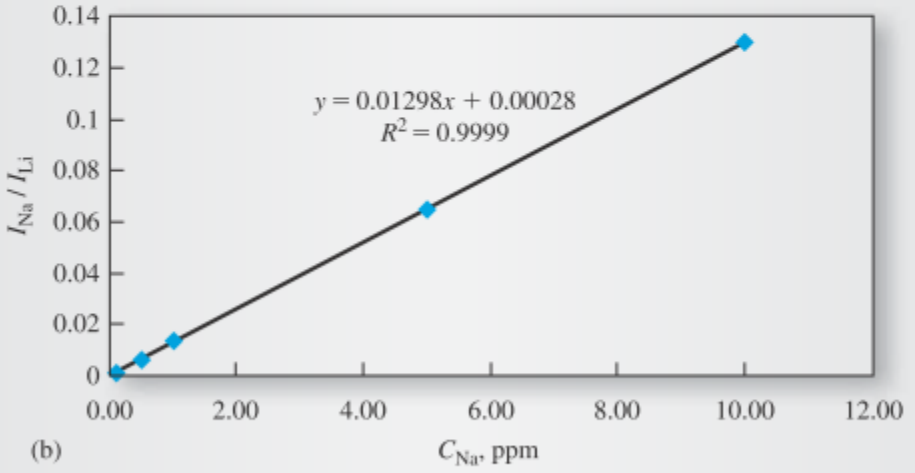


Figure 8-13 In (a) the Na flame emission intensity is plotted versus the Na concentration in ppm. The internal standard calibration curve is shown in (b), where the ratio of the Na to Li intensities is plotted versus the Na concentration.

Standard Addition Methods

*The method of standard additions is used when it is difficult or impossible to duplicate the sample matrix.

*A known amount of a standard solution of analyte is added to one portion of the sample.

*The responses before and after the addition are measured and used to obtain the analyte concentration.

*Alternatively, multiple additions are made to several portions of the sample.

*The standard additions method assumes a linear response.

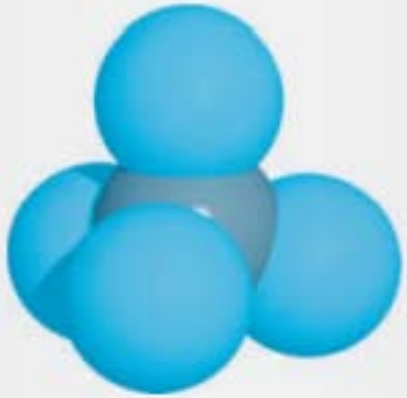
*Linearity should always be confirmed, or the multiple additions method used to check linearity.

*The method of standard additions is quite powerful so long as there is a good blank measurement so that extraneous species do not contribute to the analytical response.

* Second, the calibration curve for the analyte must be linear in the sample matrix.

EXAMPLE 8-8

The single-point standard addition method was used in the determination of phosphate by the molybdenum blue method. A 2.00-mL urine sample was treated with molybdenum blue reagents to produce a species absorbing at 820 nm, after which the sample was diluted to 100.00 mL. A 25.00-mL aliquot gave an instrument reading (absorbance) of 0.428 (solution 1). Addition of 1.00 mL of a solution containing 0.0500 mg of phosphate to a second 25.0-mL aliquot gave an absorbance of 0.517 (solution 2). Use these data to calculate the concentration of phosphate in milligrams per mL of the sample. Assume that there is a linear relationship between absorbance and concentration and that a blank measurement has been made.



Molecular model of phosphate ion (PO_4^{3-}).

Solution

The absorbance of the first solution is given by

$$A_1 = kc_u$$

where c_u is the unknown concentration of phosphate in the first solution and k is a proportionality constant. The absorbance of the second solution is given by

$$A_2 = \frac{kV_u c_u}{V_t} + \frac{kV_s c_s}{V_t}$$

where V_u is the volume of the solution of unknown phosphate concentration (25.00 mL), V_s is the volume of the standard solution of phosphate added (1.00 mL), V_t is the total volume after the addition (26.00 mL), and c_s is the concentration of the standard solution (0.500 mg mL^{-1}). If we solve the first equation for k, substitute the result into the second equation, and solve for c_u , we obtain

$$c_u = \frac{A_1 c_s V_s}{A_2 V_t - A_1 V_u} =$$

$$= \frac{0.428 \times 0.0500 \text{ mg mL}^{-1} \times 1.00 \text{ mL}}{0.517 \times 26.00 \text{ mL} - 0.428 \times 25.00 \text{ mL}} = 0.0780 \text{ mg mL}^{-1}$$

This is the concentration of the diluted sample. To obtain the concentration of the original urine sample, we need to multiply by 100.00/2.00. Thus,

$$\text{concentration of phosphate} = 0.00780 \text{ mg mL}^{-1} \times 100.00 \text{ mL}/2.00 \text{ mL}$$
$$= 0.390 \text{ mg mL}^{-1}$$

8E Figures of merit for analytical methods

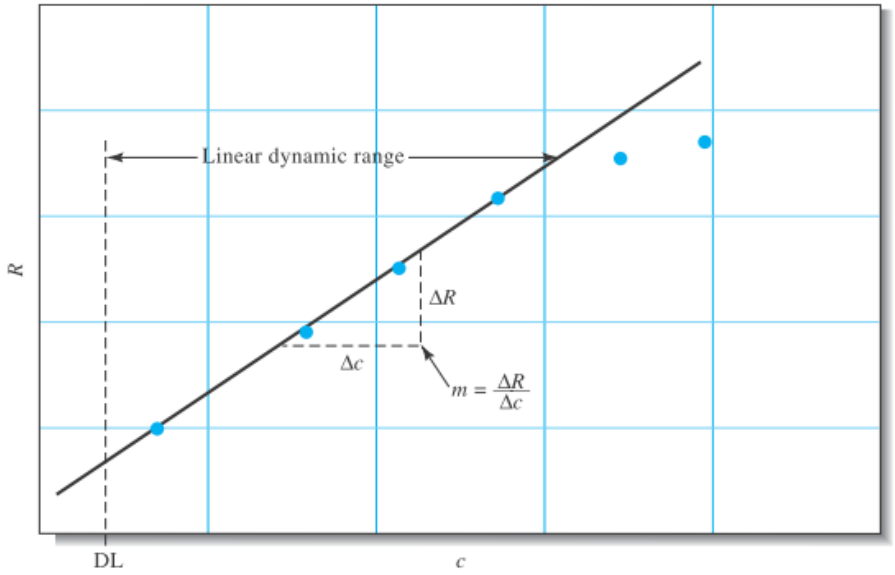
Analytical procedures are characterized by a number of figures of merit such as:

- accuracy, precision, *sensitivity, detection limit, and dynamic range*.

8E-1 Sensitivity and Detection Limit

The definition of sensitivity most often used is the *calibration sensitivity*, or the change in the response signal per unit change in analyte concentration.

The *calibration sensitivity is thus the slope of the calibration curve.* The calibration sensitivity does not indicate what concentration differences can be detected. Noise in the response signals must be taken into account in order to be quantitative about what differences can be detected. For this reason, the term **analytical sensitivity** is sometimes used**.**



The ***analytical sensitivity*** is the ratio of the calibration curve slope to the standard deviation of the analytical signal at a given analyte concentration. The analytical sensitivity is usually a strong function of concentration.

The detection limit, DL, is the smallest concentration that can be reported with a certain level of confidence.

Figure 8-14 Calibration curve of response *R versus concentration c.* The slope of the calibration curve is called the calibration sensitivity *m.* The detection limit, DL, designates the lowest concentration that can be measured at a specified confidence level.

* Every analytical technique has a detection limit.
* It is the analyte concentration that produces a response equal to k times the standard deviation of the blank σ_b

$$DL = \frac{ks_b}{m}$$

- where k is called the confidence factor and m is the calibration sensitivity. The factor k is usually chosen to be 2 or 3. A *k value of 2 corresponds to a confidence level of* 92.1%, while a *k value of 3 corresponds to a 98.3% confidence level.*

Linear Dynamic Range

* The linear dynamic range of an analytical method most often refers to the concentration range over which the analyte can be determined using a linear calibration curve .

* The lower limit is generally considered to be the detection limit.

* The upper end is usually taken as the concentration at which the analytical signal or the slope of the calibration curve deviates by a specified amount.

* Usually a deviation of 5% from linearity is considered the upper limit.

Quality Assurance of Analytical Results

Control Charts

* A control chart is a sequential plot of some quality characteristic that is important in quality assurance.

* The chart also shows the statistical limits of variation, the upper control limit (UCL) and lower control limit (LCL), that are permissible for the characteristic being measured.

$$UCL = \mu + \frac{3\sigma}{\sqrt{N}} \qquad LCL = \mu + \frac{3\sigma}{\sqrt{N}}$$

Where μ is the population mean
σ is the population standard deviation and
N, number of replicates for each sample.

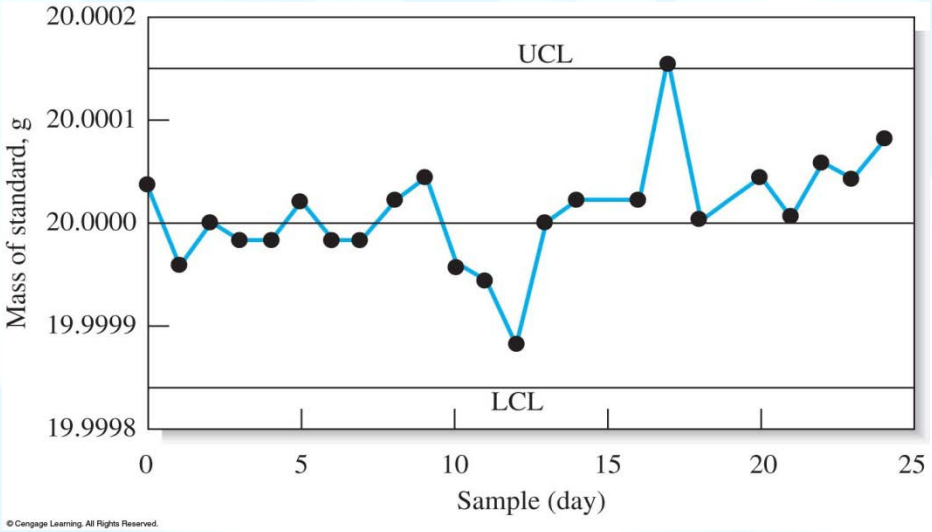


Figure 8-15 A control chart for a modern analytical balance. As long as the mean mass remains between the LCL and the UCL, the balance is said to be in statistical control.

Mass data were collected on twenty-four consecutive days for a 20.000-g standard mass certified by the National Institute of Standards and Technology. On each day, five replicate determinations were made. From independent experiments, estimates of the population mean and standard deviation were found to be μ= 20.000 g and σ=0.00012 g, respectively. For the mean of five measurements, 3x0.00012/√5 = 0.00016. Hence, the UCL value = 20.00016 g, and the LCL value = 19.99984 g.

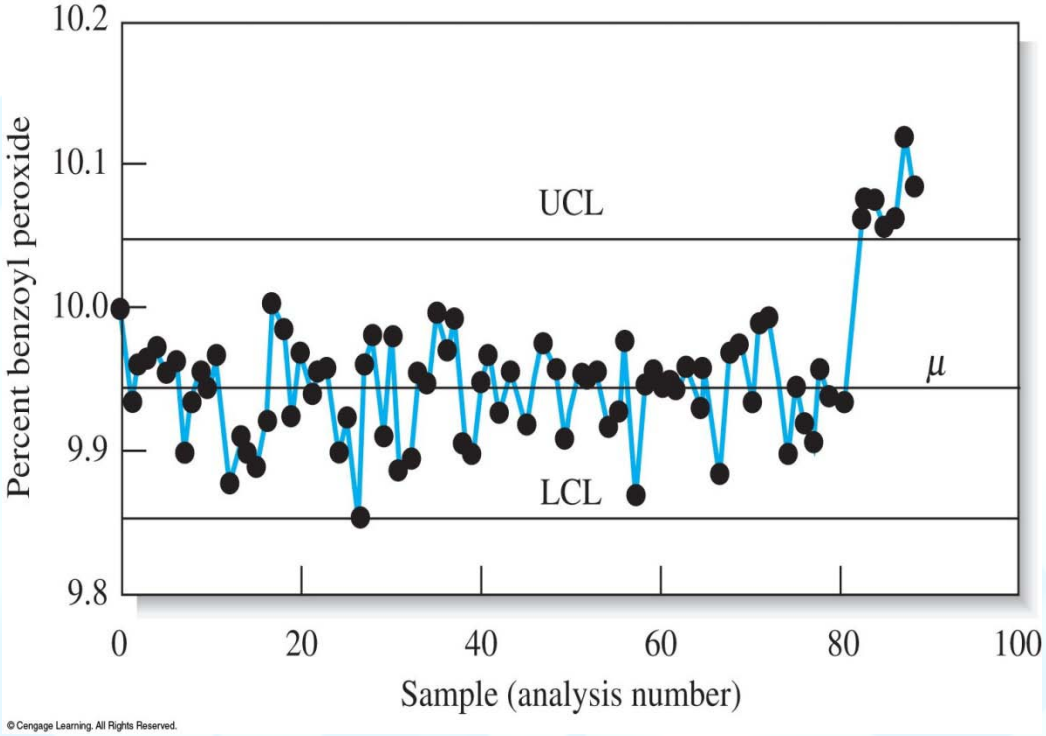


Figure 8-16 A control chart for monitoring the concentration of benzoyl peroxide in a commercial acne preparation.

The manufacturing process became out of statistical control with sample 83 and exhibited a systematic change in the mean concentration.

Validation

- Validation determines the suitability of an analysis for providing the sought-for information and can apply to samples, to methodologies, and to data.
- Validation is often done by the analyst, but it can also be done by supervisory personnel.
- There are several different ways to validate analytical methods. The most common methods include:
  - analysis of standard reference materials when available,
  - analysis by a different analytical method,
  - analysis of “spiked” samples, and
  - analysis of synthetic samples approximating the chemical composition of the test samples.
- Individual analysts and laboratories often must periodically demonstrate the validity of the methods and techniques used.
- Data validation is the final step before release of the results. This process starts with validating the samples and methods used. Then, the data are reported with statistically valid limits of uncertainty after a thorough check has been made to eliminate blunders in sampling and sample handling, mistakes in performing the analysis, errors in identifying samples, and mistakes in the calculations used.

Reporting Analytical Results

- Analytical results should be reported as the mean value and the standard deviation.
- Sometimes, the standard deviation of the mean is reported instead of that of the data
- A confidence interval for the mean, the interval and its confidence level should be explicitly reported.
- The results of various statistical tests on the data should also be reported when appropriate, as should the rejection of any outlying results along with the rejection criterion.
- Significant figures are quite important when reporting results and should be based on statistical evaluation of the data.
- Whenever possible graphical presentation should include error bars on the data points to indicate uncertainty.

Suggested Problems

8.4, 8.10, 8.13, 8.17, 8.20, 8-23