chapter 4

Evaluation of Analytical Data

J.Scott Smith

*Department of Animal Sciences and Industry,
Kansas State University,
Manhattan, KS 66506-1600, USA
e-mail: jsschem@ksu.edu*

4.1   INTRODUCTION

The field of food analysis, or any type of analysis, involves a considerable amount of time learning principles, methods, and instrument operations and perfecting various techniques. Although these areas are extremely important, much of our effort would be for naught if there were not some way for us to evaluate the data obtained from the various analytical assays. Several mathematical treatments are available that provide an idea of how well a particular assay was performed or how well we can reproduce an experiment. Fortunately, the statistics are not too involved and apply to most analytical determinations.

Whether analytical data are collected in a research laboratory or in the food industry, important decisions are made based on the data. Appropriate data collection and analysis help avoid bad decisions being made based on the data. Having a good understanding of the data and how to interpret the data (e.g., what numbers are statistically the same) are critical to good decision making. Talking with a statistician before designing experiments or testing products produced can help ensure appropriate data collection and analysis, for better decision making.

The focus in this chapter is primarily on how to evaluate replicate analyses of the same sample for accuracy and precision. In addition, considerable attention is given to the determination of best line fits for standard curve data. Keep in mind as you read and work through this chapter that there is a vast array of computer software to perform most types of data evaluation and calculations/plots.

Proper sampling and sample size are not covered in this chapter. Readers should refer to Chap. 5 and Garfield et al. [1] for sampling in general and statistical approaches to determine the appropriate sample size, and to Chap. 33, Sect. 33.4 for mycotoxin sampling.

4.2   MEASURES OF CENTRAL TENDENCY

To increase accuracy and precision, as well as to evaluate these parameters, the analysis of a sample is usually performed (repeated) several times. At least three assays are typically performed, though often the number can be much higher. Because we are not sure which value is closest to the true value, we determine the mean (or average) using all the values obtained and report the results of the **mean**. The mean is designated by the symbol $\bar{x}$ and calculated according to the equation below:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\Sigma x_i}{n} \qquad (4.1)$$

where:

$$\bar{x} = \text{mean}$$

$$x_1, x_2, \text{etc.} = \text{individually measured values } (x_i)$$

$$n = \text{number of measurements}$$

For example, suppose we measured a sample of uncooked hamburger for percent moisture content four times and obtained the following results: 64.53 %, 64.45 %, 65.10 %, and 64.78 %:

$$\bar{x} = \frac{64.53 + 64.45 + 65.10 + 64.78}{4} = 64.72\% \qquad (4.2)$$

Thus, the result would be reported as 64.72 % moisture. When we report the mean value, we are indicating that this is the best experimental estimate of the value. We are not saying anything about how accurate or true this value is. Some of the individual values may be closer to the true value, but there is no way to make that determination, so we report only the mean.

Another determination that can be used is the **median**, which is the midpoint or middle number within a group of numbers. Basically, half of the experimental values will be less than the median and half will be greater. The median is not used often, because the mean is such a superior experimental estimator.

4.3   RELIABILITY OF ANALYSIS

Returning to our previous example, recall that we obtained a mean value for moisture. However, we did not have any indication of how repeatable the tests were or how close our results were to the true value. The next several sections will deal with these questions and some of the relatively simple ways to calculate the answers. More thorough coverage of these areas is found in references [2–4].

4.3.1   Accuracy and Precision

One of the most confusing aspects of data analysis for students is grasping the concepts of accuracy and precision. These terms are commonly used interchangeably in society, which only adds to this confusion. If we consider the purpose of the analysis, then these terms become much clearer. If we look at our experiments, we know that the first data obtained are the individual results and a mean value ($\bar{x}$). The next questions should be: "How close were our individual measurements?" and "How close were they to the true value?" Both questions involve accuracy and precision. Now, let us turn our attention to these terms.

Accuracy refers to how close a particular measure is to the true or correct value. In the moisture analysis for hamburger, recall that we obtained a mean of 64.72%. Let us say the true moisture value was actually 65.05%. By comparing these two numbers, you could probably make a guess that your results were fairly accurate because they were close to the correct value. (The calculations of accuracy will be discussed later.)

The problem in determining accuracy is that most of the time we are not sure what the true value is. For certain types of materials, we can purchase known samples from, for example, the National Institute of Standards and Technology and check our assays against these samples. Only then can we have an indication of the accuracy of the testing procedures. Another approach is to compare our results with those of other labs to determine how well they agree, assuming the other labs are accurate.

A term that is much easier to deal with and determine is **precision**. This parameter is a measure of how reproducible or how close replicate measurements become. If repetitive testing yields similar results, then we would say the precision of that test was good. From a true statistical view, the precision often is called **error**, when we are actually looking at experimental **variation**. So, the concepts of precision, error, and variation are closely related.

The difference between precision and accuracy can be illustrated best with Fig. 4.1. Imagine shooting a rifle at a target that represents experimental values. The bull's eye would be the true value, and where the bullets hit would represent the individual experimental values. As you can see in Fig. 4.1a, the values can be tightly spaced (good precision) and close to the bull's eye (good accuracy), or, in some cases, there can be situations with good precision but poor accuracy (Fig. 4.1b). The worst situation, as illustrated in Fig. 4.1d, is when both the accuracy and precision are poor. In this case, because of errors or variation in the determination, interpretation of the results becomes very difficult. Later, the practical aspects of the various types of error will be discussed.

When evaluating data, several tests are commonly used to give some appreciation of how much the experimental values would vary if we were to repeat the test (indicators of precision). An easy way to look at the variation or scattering is to report the range of the experimental values. The **range** is simply the difference between the largest and smallest observation. This measurement is not too useful and thus is seldom used in evaluating data.

Probably the best and most commonly used statistical evaluation of the precision of analytical data is the standard deviation. The **standard deviation** measures the spread of the experimental values and gives a good indication of how close the values are to each other. When evaluating the standard deviation, one has to remember that we are never able to analyze the entire food product. That would be difficult, if not impossible, and very time consuming. Thus, the calculations we use are only estimates of the unknown true value.

If we have many samples, then the standard deviation is designated by the Greek letter sigma (σ). It is calculated according to Eq. 4.3, assuming all of the food product was evaluated (which would be an infinite amount of assays):

$$\sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{n}} \qquad (4.3)$$

where:

σ = standard deviation
xi = individual sample values
μ = true mean
n = total population of samples

Because we do not know the value for the true mean, the equation becomes somewhat simplified so that we can use it with real data. In this case, we now call the σ term the standard deviation of the sample and designate it by SD or σ. It is determined according to the calculation in Eq. 4.4, where $\bar{x}$ replaces the true mean term μ and n represents the number of samples:

a   **b**   **c**   **d**
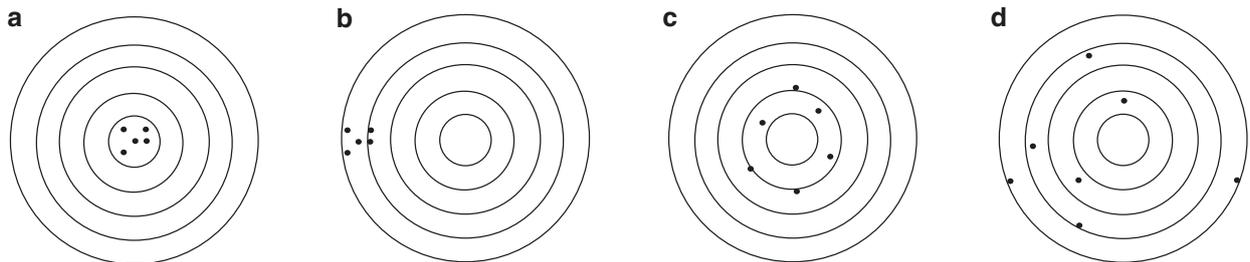
4.1 figure Comparison of accuracy and precision: (**a**) good accuracy and good precision, (**b**) good precision and poor accuracy, (**c**) good accuracy and poor precision, and (**d**) poor accuracy and poor precision

$$SD = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}} \qquad (4.4)$$

If the number of replicate determinations is small (about 30 or less), which is common with most assays, the n is replaced by the $n-1$ term, and Eq. 4.5 is used. Unless you know otherwise, Eq. 4.5 is always used in calculating the standard deviation of a group of assays:

$$SD = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} \qquad (4.5)$$

Depending on which of the equations above is used, the standard deviation may be reported as SDn or σn and SDn_{-1} or σn_{-1}. (Different brands of software and scientific calculators sometimes use different labels for the keys, so one must be careful.) Table 4.1 shows an example of the determination of standard deviation. The sample results would be reported to average 64.72 % moisture with a standard deviation of 0.293.

Once we have a mean and standard deviation, we must next determine how to interpret these numbers. One easy way to get a feel for the standard deviation is to calculate what is called the **coefficient of variation** (CV), also known as the **relative standard deviation**. This calculation is shown below for our example of the moisture determination of uncooked hamburger:

$$\% \text{Coefficient of variation} (\%CV) = \frac{SD}{\bar{x}} \times 100 \qquad (4.6)$$

$$\% CV = \frac{0.293}{64.72} \times 100 = 0.453\% \qquad (4.7)$$

The CV tells us that our standard deviation is only 0.453 % as large as the mean. For our example, that number is small, which indicates a high level of precision or reproducibility of the replicates. As a rule, a CV below 5 % is considered acceptable, although it depends on the type of analysis.

Another way to evaluate the meaning of the standard deviation is to examine its origin in statistical theory. Many populations (in our case, sample values or means) that exist in nature are said to have a normal distribution. If we were to measure an infinite number of samples, we would get a distribution similar to that represented by Fig. 4.2. In a population with a **normal distribution**, 68 % of those values would be within ±1 standard deviation from the mean, 95 % would be within ±2 standard deviations, and 99.7 % would be within ±3 standard deviations. In other words, there is a probability of less than 1 % that a sample in a population would fall outside ±3 standard deviations from the mean value.

Another way of understanding the normal distribution curve is to realize that the probability of finding the true mean is within certain confidence intervals as defined by the standard deviation. For large numbers of samples, we can determine the **confidence limit** or **interval** around the mean using the statistical parameter called the **Z value**. We do this calculation by first looking up the Z value from statistical tables once we have decided the desired degree of certainty. Some Z values are listed in Table 4.2.

The confidence limit (or interval) for our moisture data, assuming a 95 % probability, is calculated according to Eq. 4.8. Since this calculation is not valid for

4.2 figure A normal distribution curve for a population or a group of analyses

4.1 table Determination of the standard deviation of percent moisture in uncooked hamburger

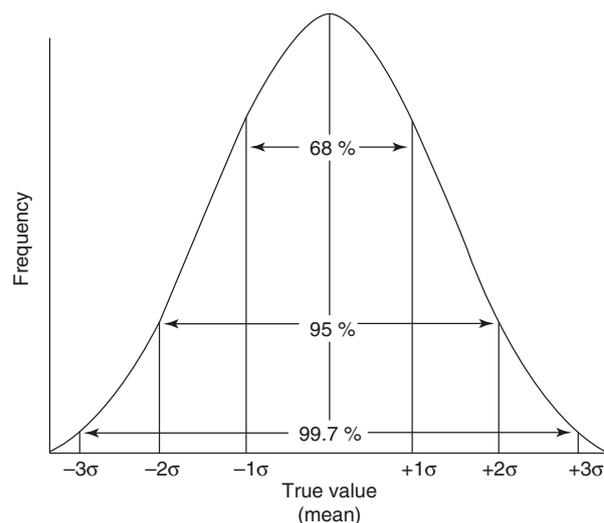| Measurement	Observed % moisture	Deviation from the mean $(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	64.53	−0.19	0.0361
2	64.45	−0.27	0.0729
3	65.10	+0.38	0.1444
4	64.78	+0.06	0.0036
	$\Sigma x_i = 258.86$		$\Sigma(x_i - \bar{x})^2 = 0.257$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{258.86}{4} = 64.72$$

$$SD = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{0.257}{3}} = 0.2927$$

4.2 table Values for Z for checking both upper and lower levels

Degree of certainty (confidence) (%)	Z value
80	1.29
90	1.64
95	1.96
99	2.58
99.9	3.29

small numbers, assume we ran 25 samples instead of four:

Confidence interval (CI)

$$= \bar{x} \pm Z \text{ value} \times \frac{\text{standard deviation (SD)}}{\sqrt{n}} \quad (4.8)$$

$$\text{CI (at 95\%)} = 64.72 \pm 1.96 \times \frac{0.2927}{\sqrt{25}}$$

$$= 64.72 \pm 0.115\% \quad (4.9)$$

Because our example had only four values for the moisture levels, the confidence interval should be calculated using statistical t-tables. In this case, we have to look up the t value from Table 4.3 based on the **degrees of freedom**, which is the sample size minus one $(n-1)$, and the desired level of confidence.

The calculation for our moisture example with four samples (n) and three degrees of freedom $(n-1)$ is given below:

$$\text{CI} = \bar{x} \pm t \text{ value} \times \frac{\text{standard deviation (SD)}}{\sqrt{n}} \quad (4.10)$$

$$\text{CI (at 95\%)} = 64.72 \pm 3.18 \times \frac{0.2927}{\sqrt{4}}$$

$$= 64.72 \pm 0.465\% \quad (4.11)$$

To interpret this number, we can say that, with 95 % confidence, the true mean for our moisture will fall within 64.72 ± 0.465 % or between 65.185 % and 64.255 %.

The expression $SD/\sqrt{n}$ is often reported as the **standard error of the mean**. It is then left to the reader to calculate the confidence interval based on the desired level of certainty.

Other quick tests of precision used are the relative deviation from the mean and the relative average deviation from the mean. The **relative deviation from the mean** is useful when only two replicates have been performed. It is calculated according to Eq. 4.12, with values below 2 % considered acceptable:

4.3
table

Values of *t* for various levels of probability[a]

| Degrees of freedom (n− 1) | Levels of certainty | | |
	95 %	99 %	99.9 %
1	12.7	63.7	636
2	4.30	9.93	31.60
3	3.18	5.84	12.90
4	2.78	4.60	8.61
5	2.57	4.03	6.86
6	2.45	3.71	5.96
7	2.36	3.50	5.40
8	2.31	3.56	5.04
9	2.26	3.25	4.78
10	2.23	3.17	4.59

[a]More extensive t-tables can be found in statistics books

$$\text{Relative deviation from the mean} = \frac{x_i - \bar{x}}{\bar{x}} \times 100 \quad (4.12)$$

where:

x_i = individual sample value
$\bar{x}$ = mean

If there are several experimental values, then the **relative average deviation from the mean** becomes a useful indicator of precision. It is calculated similarly to the relative deviation from the mean, except the average deviation is used instead of the individual deviation. It is calculated according to Eq. 4.13:

Relative average deviation from the mean

$$= \frac{\frac{\Sigma|x_i - \bar{x}|}{n}}{\bar{x}} \times 1000$$

$$= \text{parts per thousand} \quad (4.13)$$

Using the moisture values discussed in Table 4.1, the $x_i - \bar{x}$ terms for each determination are −0.19, −0.27, +0.38, and +0.06. Thus, the calculation becomes:

$$\text{Rel. avg. dev.} = \frac{\frac{0.19 + 0.27 + 0.38 + 0.06}{4}}{64.72} \times 1000$$

$$= \frac{0.225}{64.72} \times 1000$$

$$= 3.47 \text{ parts per thousand} \quad (4.14)$$

Up to now, our discussions of calculations have involved ways to evaluate precision. If the true value is not known, we can calculate only precision. A low degree of precision would make it difficult to predict a realistic value for the sample.

However, we may occasionally have a sample for which we know the true value and can compare our results with the known value. In this case, we can calculate the error for our test, compare it to the known value, and determine the accuracy. One term that can be calculated is the **absolute error**, which is simply the difference between the experimental value and the true value:

$$\text{Absolute error} = E_{abs} = x - T \quad (4.15)$$

where:

x = experimentally determined value
T = true value

The absolute error term can have either a positive or negative value. If the experimentally determined value is from several replicates, then the mean (0) would be substituted for the x term. This is not a good test for error, because the value is not related to the

magnitude of the true value. A more useful measurement of error is **relative error**:

$$\text{Relative error} = E_{\text{rel}} = \frac{E_{\text{abs}}}{T} = \frac{x - T}{T} \qquad (4.16)$$

The results are reported as a negative or positive value, which represents a fraction of the true value.

If desired, the relative error can be expressed as percent relative error by multiplying by 100 %. Then the relationship becomes the following, where x can be either an individual determination or the mean (0) of several determinations:

$$\% E_{\text{rel}} = \frac{E_{\text{abs}}}{T} \times 100 = \frac{x - T}{T} \times 100 \qquad (4.17)$$

Using the data for the percent moisture of uncooked hamburger, suppose the true value of the sample is 65.05 %. The percent relative error is calculated using our mean value of 64.72 % and Eq. 4.17:

$$\% E_{\text{rel}} = \frac{\overline{x} - T}{T} \times 100 = \frac{64.72 - 65.05}{65.05} \times 100$$
$$= -0.507\% \qquad (4.18)$$

Note that we keep the negative value, which indicates the direction of our error, that is, our results were 0.507 % lower than the true value.

4.3.2 Sources of Errors [3]

As you may recall from the discussions of accuracy and precision, error (variation) can be quite important in analytical determinations. Although we strive to obtain correct results, it is unreasonable to expect an analytical technique to be entirely free of error. The best we can hope for is that the variation is small and, if possible, at least consistent. As long as we know about the error, the analytical method often will be satisfactory. There are several sources of error, which can be classified as: systematic error (determinate), random error (indeterminate), and gross error or blunders. Again, note that error and variation are used interchangeably in this section and essentially have the same meaning for these discussions.

Systematic or **determinate error** produces results that consistently deviate from the expected value in one direction or the other. As illustrated in Fig. 4.1b, the results are spaced closely together, but they are consistently off the target. Identifying the source of this serious type of error can be difficult and time consuming, because it often involves inaccurate instruments or measuring devices. For example, a pipette that consistently delivers the wrong volume of reagent will produce a high degree of precision yet inaccurate results. Sometimes impure chemicals or the analytical method itself is the cause. Generally, we can overcome systematic errors by proper calibration of instruments,

running blank determinations, or using a different analytical method.

Random or **indeterminate errors** are always present in any analytical measurement. This type of error is due to our natural limitations in measuring a particular system. These errors fluctuate in a random fashion and are essentially unavoidable. For example, reading an analytical balance, judging the endpoint change in a titration, and using a pipette all contribute to random error. Background instrument noise, which is always present to some extent, is a factor in random error. Both positive and negative errors are equally possible. Although this type of error is difficult to avoid, fortunately it is usually small.

Blunders are easy to eliminate, since they are so obvious. The experimental data are usually scattered, and the results are not close to an expected value. This type of error is a result of using the wrong reagent or instrument or of sloppy technique. Some people have called this type of error the "Monday morning syndrome" error. Fortunately, blunders are easily identified and corrected.

4.3.3 Specificity

Specificity of a particular analytical method means that it detects only the component of interest. Analytical methods can be very specific for a certain food component or, in many cases, can analyze a broad spectrum of components. Quite often, it is desirable for the method to be somewhat broad in its detection. For example, the determination of food lipid (fat) is actually the crude analysis of any compound that is soluble in an organic solvent. Some of these compounds are glycerides, phospholipids, carotenes, and free fatty acids. Since we are not concerned about each individual compound when considering the crude fat content of food, it is desirable that the method be broad in scope. On the other hand, determining the lactose content of ice cream would require a specific method. Because ice cream contains other types of simple sugars, without a specific method, we would overestimate the amount of lactose present.

There are no hard rules for what specificity is required. Each situation is different and depends on the desired results and type of assay used. However, it is something to keep in mind as the various analytical techniques are discussed.

4.3.4 Sensitivity and Limit of Detection [5]

Although often used interchangeably, the terms sensitivity and limit of detection should not be confused. They have different meanings, yet they are closely related. **Sensitivity** relates to the magnitude of change of a measuring device (instrument) with changes in compound concentration. It is an indicator of how lit-

tle change can be made in the unknown material before we notice a difference on a needle gauge or a digital readout. We are all familiar with the process of tuning in a radio station on our stereo and know how, at some point, once the station is tuned in, we can move the dial without disturbing the reception. This is sensitivity. In many situations, we can adjust the sensitivity of an assay to fit our needs, that is, whether we desire more or less sensitivity. We even may desire a lower sensitivity so that samples with widely varying concentration can be analyzed at the same time.

Limit of detection (LOD), in contrast to sensitivity, is the lowest possible amount that we can detect with some degree of confidence (or statistical significance). With every assay, there is a lower limit at which point we are not sure if something is present or not. Obviously, the best choice would be to concentrate the sample so we are not working close to the detection limit. However, this may not be possible, and we may need to know the LOD so we can work away from that limit.

There are several ways to measure the LOD, depending on the apparatus that is used. If we are using something like a spectrophotometer, gas chromatograph, or high-performance liquid chromatography (HPLC), the LOD often is reached when the signal to noise ratio is 3 or greater [5]. In other words, when the sample gives a value that is three times the magnitude of the noise detection, the instrument is at the lowest limit possible. Noise is the random signal fluctuation that occurs with any instrument.

A more general way to define the limit of detection is to approach the problem from a statistical viewpoint, in which the variation between samples is considered. A common mathematical definition of limit of detection is given below [3]:

$$X_{LD} = X_{Blk} + (3 \times SD_{Blk}) \qquad (4.19)$$

where:

X_{LD} = minimum detectable concentration
X_{Blk} = signal of a blank
SD_{Blk} = standard deviation of the blank readings

In this equation, the variation of the blank values (or noise, if we are talking about instruments) determines the detection limit. High variability in the blank values increases the limit of detection.

Another method that encompasses the entire assay method is the **method detection limit** (MDL). According to the US Environmental Protection Agency (EPA) [6], the MDL is defined as "the minimum concentration of a substance that can be measured and reported with 99 % confidence that the analyte concentration is greater than zero and is determined from

analysis of a sample in a given matrix containing the analyte." What differentiates the MDL from the LOD is that it includes the entire assay and various sample types thus correcting for variability throughout. The MDL is calculated based on values of samples within the assay matrix and thus is considered a more rigorous performance test. The procedures on how to set up the MDL are explained in Appendix B of Part 136 (40 CFR, Vol 22) of the EPA regulations on environmental testing.

Though the LOD or MDL is often sufficient to characterize an assay, a further evaluation to check is the **limit of quantitation** (LOQ). In this determination data are collected similar to the LOD except the value is determined as $X_{Blk} + (10 \times SD_{Blk})$ instead of $(X_{Blk} + 3 \times SD_{Blk})$.

4.3.5 Quality Control Measures [1–3]

Quality control/assurance is desirable to evaluate analysis performance of a method or process. To explain how analytical data and control charts can be used in the food industry for **statistical process control**, this section will briefly describe quality control from the perspective of monitoring a specific process in making a food product (e.g., drying of a product, thereby affecting final moisture content). If the process is well defined and has known variability, the analytical data gathered can be evaluated over time. This provides set control points to determine if the process is performing as intended. Since all processes are susceptible to changes or drift, a decision can be made to adjust the process.

The best way to evaluate quality control is by control charting. This entails sequential plotting of the mean observations (e.g., moisture content) obtained from the analysis along with a target value. The standard deviation then is used to determine acceptable limits at the 95 % or 99 % confidence level, and at what point the data are outside the range of acceptable values. Often the acceptable limits are set as two standard deviations on either side of the mean, with the action limits set at three standard deviations. The charts and limits are used to determine if variation has occurred that is outside the normal variation for the process. If this occurs, there is a need to determine the root cause of the variation and put in place corrective and preventive actions to further improve the process.

Two common types of **control charts** used are the Shewhart and CuSum charts described by Ellison et al. [2]. The CuSum chart is more involved and is better at highlighting small changes in the mean value. The Shewhart chart (Fig. 4.3) entails plots of the target value mean and both upper and lower limits for each measurement. An upper and lower warning limit and action limit are determined and added to the plot. The warning limit shows that the
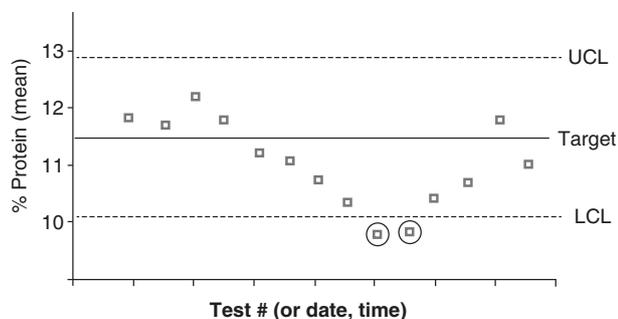
4.3 figure An example of a Shewhart control chart for protein analysis. The upper control limit (UCL) and lower control limit (LCL) are predetermined. Values which fall out of the limits (the circled values) indicate that the assay requires action and needs to be corrected or adjusted

measurements may be moving out of the desirable limit (e.g., upward trend). The action limit indicates that the measurements are past the acceptable limit so the process needs to be evaluated for the causes of the drift. Examples of the calculations and charts are provided in references [2, 3].

4.4 CURVE FITTING: REGRESSION ANALYSIS [2–4]

Curve fitting is a generic term used to describe the relationship and evaluation between two variables. Most scientific fields use curve fitting procedures to evaluate the relationship of two variables. Thus, curve fitting or curvilinear analysis of data is a vast area as evidenced by the volumes of material describing these procedures. In analytical determinations, we are usually concerned with only a small segment of curvilinear analysis, the standard curve, or regression line.

A **standard curve** or **calibration curve** is used to determine unknown concentrations based on a method that gives some type of measurable response that is proportional to a known amount of standard. It typically involves making a group of known standards in increasing concentration and then recording the particular measured analytical parameter (e.g., absorbance, area of a chromatography peak, etc.). What results when we graph the paired x and y values is a scatter plot of points that can be joined together to form a straight line relating concentration to observed response. Once we know how the observed values change with concentration, it is fairly easy to estimate the concentration of an unknown by interpolation from the standard curve.

As you read through the next three sections, keep in mind that not all correlations of observed values to

standard concentrations are linear (but most are). There are many examples of nonlinear curves, such as antibody binding, toxicity evaluations, and exponential growth and decay. Fortunately, with the vast array of computer software available today, it is relatively easy to analyze any group of data.

4.4.1 Linear Regression [2–4]

So how do we set up a standard curve once the data have been collected? First, a decision must be made regarding onto which axis to plot the paired sets of data. Traditionally, the concentration of the standards is represented on the x-axis, and the observed readings are on the y-axis. However, this protocol is used for reasons other than convention. The x-axis data are called the **independent variable** and are assumed to be essentially free of error, while the y-axis data (the **dependent variable**) may have error associated with them. This assumption may not be true because error could be incorporated as the standards are made. With modern-day instruments, the error can be very small. Although arguments can be made for making the y-axis data concentration, for all practical purposes, the end result is essentially the same. Unless there are some unusual data, the *concentration should be associated with the x-axis and the measured values with the y-axis*.

Figure 4.4 illustrates a typical standard curve used in the determination of caffeine in various foods. Caffeine is analyzed readily in foods by using HPLC coupled with an ultraviolet detector set at 272 nm. The area under the caffeine peak at 272 nm is directly proportional to the concentration. When an unknown sample (e.g., coffee) is run on the HPLC, a peak area is

Conc (x)	Peak Area (y)
5	488
25	2291
50	4795
75	6780
100	9048
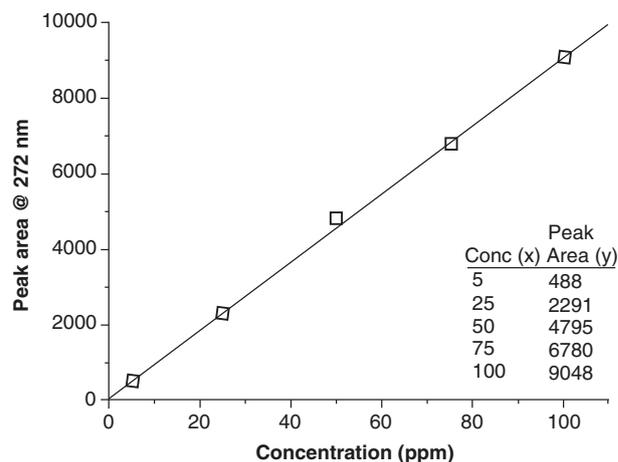
4.4 figure A typical standard curve plot showing the data points and the generated best fit line. The data used to plot the curve are presented on the graph

obtained that can be related back to the sample using the standard curve.

The plot in Fig. 4.4 shows all the data points and a straight line that appears to pass through most of the points. The line almost passes through the origin, which makes sense because zero concentration should produce no signal at 272 nm. However, the line is not perfectly straight (and never is) and does not quite pass through the origin.

To determine the caffeine concentration in a sample that gave an area of say 4,000, we could interpolate to the line and then draw a line down to the x-axis. Following a line to the x-axis (concentration), we can estimate the solution to be at about 42–43 ppm of caffeine.

We can mathematically determine the best fit of the line by using **linear regression**. Keep in mind the equation for a straight line, which is $y = ax + b$, where a is the slope and b is the y-intercept. To determine the slope and y-intercept, the regression equations shown below are used. We determine a and b and thus, for any value of y (measured), we can determine the concentration (x):

$$\text{slope } a = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} \quad (4.20)$$

$$y - \text{intercept } b = \bar{y} - a\bar{x} \quad (4.21)$$

where:

x_i and y_i = individual values

$\bar{x}$ and $\bar{y}$ = means of the individual values

Low-cost calculators and computer spreadsheet software can readily calculate regression equations, so no attempt is made to go through the mathematics in the formulas.

The formulas give what is known as the line of regression of y on x, which assumes that the error occurs in the y direction. The regression line represents the average relationship between all the data points and thus is a balanced line. These equations also assume that the straight-line fit does not have to go through the origin, which at first does not make much sense. However, there are often background interferences, so that even at zero concentration, a weak signal may be observed. In most situations, calculating the origin as going through zero will yield the same results.

Using the data from Fig. 4.4, calculate the concentration of caffeine in the unknown and compare with the graphing method. As you recall, the unknown had an area at 272 nm of 4,000. Linear regression analysis of the standard curve data gave the y-intercept (b) as 90.727 and the slope (a) as 89.994 ($r^2 = 0.9989$):

$$y = ax + b \quad (4.22)$$

or

$$x = \frac{y - b}{a} \quad (4.23)$$

$$x\,(\text{conc}) = \frac{4000 - 90.727}{89.994} = 43.4393 \text{ ppm caffeine} \quad (4.24)$$

The agreement is fairly close when comparing the calculated value to that estimated from the graph. Using high-quality graph paper with many lines could give us a line very close to the calculated one. However, as we will see in the next section, additional information can be obtained about the nature of the line when using computer software or calculators.

4.4.2 Correlation Coefficient

In observing any type of correlation, including linear ones, questions always surface concerning how to draw the line through the data points and how well the data fit to the straight line. The first thing that should be done with any group of data is to plot it to see if the points fit a straight line. By just looking at the plotted data, it is fairly easy to make a judgment on the linearity of the line. We also can pick out regions on the line where a linear relationship does not exist. The figures below illustrate differences in standard curves; Fig. 4.5a shows a good correlation of the data and Fig. 4.5b shows a poor correlation. In both cases, we can draw a straight line through the data points. Both curves yield the same straight line, but the precision is poorer for the latter.

There are other possibilities when working with standard curves. Figure 4.6a shows a good correlation between x and y, but in the negative direction, and Fig. 4.6b illustrates data that have no correlation at all.

The **correlation coefficient** defines how well the data fit to a straight line. For a standard curve, the ideal situation would be that all data points lie perfectly on a
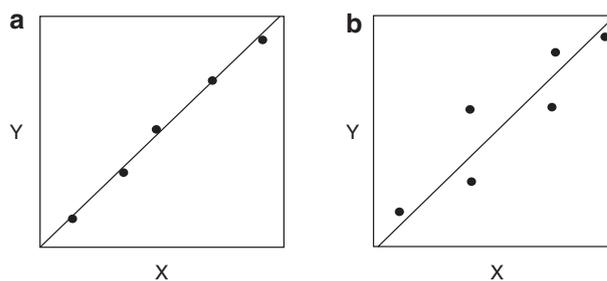
4.5 figure Examples of standard curves showing the relationship between the x and y variables when there are (**a**) a high amount of correlation and (**b**) a lower amount of correlation. Both lines have the same equation
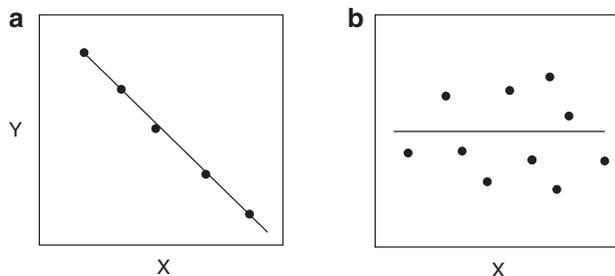
4.6 figure

Examples of standard curves showing the relationship between the x and y variables when there are (**a**) a high amount of negative correlation and (**b**) no correlation between x and y values

straight line. However, this is never the case, because errors are introduced in making standards and measuring the physical values (observations).

The correlation coefficient and coefficient of determination are defined below. Essentially all spreadsheet and plotting software will calculate the values automatically:

correlation coefficient =

$$r = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\Sigma (x_i - \bar{x})^2\right]\left[\Sigma (y_i - \bar{y})^2\right]}} \quad (4.25)$$

For our example of the caffeine standard curve from Fig. 4.3, $r = 0.99943$ (values are usually reported to at least four significant figures).

For standard curves, we want the value of r as close to $+1.0000$ or -1.000 as possible, because this value is a perfect correlation (perfect straight line). Generally, in analytical work, the r should be 0.9970 or better (this does not apply to biological studies).

The **coefficient of determination** (r^2) is used often because it gives a better perception of the straight line even though it does not indicate the direction of the correlation. The $r2$ for the example presented above is 0.99886, which represents the proportion of the variance of absorbance (y) that can be attributed to its linear regression on concentration (x). This means that about 0.114% of the straight-line variation ($1.0000 - 0.99886 = 0.00114 \times 100\% = 0.114\%$) does not vary with changes in x and y and thus is due to indeterminate variation. A small amount of variation is expected normally.

4.4.3 Errors in Regression Lines

While the correlation coefficient tells us something about the error or variation in linear curve fits, it does not always give the complete picture. Also, neither linear regression nor correlation coefficient will indicate that a particular set of data have a linear relationship. They only provide an estimate of the fit assuming the

line is a linear one. As indicated before, plotting the data is critical when looking at how the data fit on the curve (actually, a line). One parameter that is used often is the **y-residuals**, which are simply the differences between the observed values and the calculated or computed values (from the regression line). Advanced computer graphics software can actually plot the residuals for each data point as a function of concentration. However, plotting the residuals is usually not necessary because data that do not fit on the line are usually quite obvious. If the residuals are large for the entire curve, then the entire method needs to be evaluated carefully. However, the presence of one point that is obviously off the line while the rest of the points fit very well probably indicates an improperly made standard.

One way to reduce the amount of error is to include more replicates of the data such as repeating the observations with a new set of standards. The replicate x and y values can be entered into the calculator or spreadsheet as separate points for the regression and coefficient determinations. Another, probably more desirable, option is to expand the concentrations at which the readings are taken. Collecting observations at more data points (concentrations) will produce a better standard curve. However, increasing the data beyond seven or eight points usually is not beneficial.

Plotting **confidence intervals**, or **bands** or **limits**, on the standard curve along with the regression line is another way to gain insight into the reliability of the standard curve. Confidence bands define the statistical uncertainty of the regression line at a chosen probability (such as 95%) using the t-statistic and the calculated standard deviation of the fit. In some aspects, the confidence bands on the standard curve are similar to the confidence interval discussed in Sect. 4.3.1. However, in this case we are looking at a line rather than a confidence interval around a mean. Figure 4.7 shows the caffeine data from the standard curve presented before, except some of the numbers have been modified to enhance the confidence bands. The confidence bands (dashed lines) consist of both an upper limit and a lower limit that define the variation of the y-axis value. The upper and lower bands are narrowest at the center of the curve and get wider as the curve moves to the higher or lower standard concentrations.
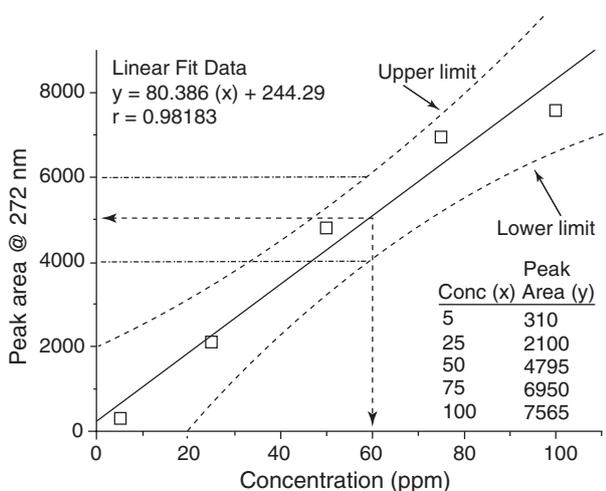
Looking at Fig. 4.7 again, note that the confidence bands show what amount of variation we expect in a peak area at a particular concentration. At 60 ppm concentration, by going up from the x-axis to the bands and interpolating to the y-axis, we see that with our data the 95% confidence interval of the observed peak area will be 4,000–6,000. In this case, the variation is large and would not be acceptable as a standard curve and is presented here only for illustration purposes.

Error bars also can be used to show the variation of y at each data point. Several types of error or variation statistics can be used such as standard error,

standard deviation, or percentage of data (i.e., 5 %). Any of these methods give a visual indication of experimental variation.
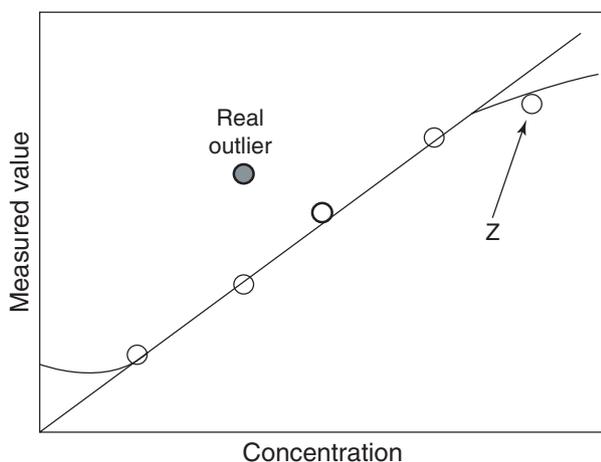
Even with good standard curve data, problems can arise if the standard curve is not used properly. One common mistake is to extrapolate beyond the data points used to construct the curve. Figure 4.8 illustrates some of the possible problems that might occur when extrapolation is used. As shown in Fig. 4.8, the curve or line may not be linear outside the area where the data were collected. This can occur in the region close to the origin or especially at the higher concentration level.

Usually a standard curve will go through the origin, but in some situations it may actually tail off as

zero concentration is approached. At the other end of the curve, at higher concentrations, it is fairly common for a plateau to be reached where the measured parameter does not change much with an increase in concentration. Care must be used at the upper limit of the curve to ensure that data for unknowns are not collected outside of the curve standards. Point Z on Fig. 4.7 should be evaluated carefully to determine if the point is an outlier or if the curve is actually tailing off. Collecting several sets of data at even higher concentrations should clarify this. Regardless, the unknowns should be measured only in the region of the curve that is linear.

4.5   REPORTING RESULTS

In dealing with experimental results, we are always confronted with reporting data in a way that indicates the sensitivity and precision of the assay. Ideally, we do not want to overstate or understate the sensitivity of the assay, and thus we strive to report a meaningful value, be it a mean, standard deviation, or some other number. The next three sections discuss how we can evaluate experimental values so as to be precise when reporting results.

4.5.1   Significant Figures

The term **significant figure** is used rather loosely to describe some judgment of the number of reportable digits in a result. Often, the judgment is not soundly based, and meaningful digits are lost or meaningless digits are retained. Exact rules are provided below to help determine the number of significant figures to report. However, it is important to keep some flexibility when working with significant figures.

Proper use of significant figures is meant to give an indication of the sensitivity and reliability of the analytical method. Thus, reported values should contain only significant figures. A value is made up of significant figures when it contains all digits known to be true and one last digit that is in doubt. For example, a value reported as 64.72 contains four significant figures, of which three digits are certain (64.7) and the last digit is uncertain. Thus, the 2 is somewhat uncertain and could be either 1 or 3. As a rule, numbers that are presented in a value represent the significant figures, regardless of the position of any decimal points. This also is true for values containing zeros, provided they are bounded on either side by a number. For example, 64.72, 6.472, 0.6472, and 6.407 all contain four significant figures. Note that the zero to the left of the decimal point is used only to indicate that there are no numbers above 1. We could have reported the value as .6472, but using the zero is better, since we know that a number was not inadvertently left off our value.

4.7 figure   A standard curve graph showing the confidence bands. The data used to plot the graph are presented on the graph as are the equation of the line and the correlation coefficient

Linear Fit Data
y = 80.386 (x) + 244.29
r = 0.98183

Conc (x)	Peak Area (y)
5	310
25	2100
50	4795
75	6950
100	7565

4.8 figure   A standard curve plot showing possible deviations in the curve in the upper and lower limits

Special considerations are necessary for zeros that may or may not be significant:

1. Zeros after a decimal point are always significant figures. For example, 64.720 and 64.700 both contain five significant figures.
2. Zeros before a decimal point with no other preceding digits are not significant. As indicated before, 0.6472 contains four significant figures.
3. Zeros after a decimal point are not significant if there are no digits before the decimal point. For example, 0.0072 has no digits before the decimal point; thus, this value contains two significant figures. In contrast, the value 1.0072 contains five significant figures.
4. Final zeros in a number are not significant unless indicated otherwise. Thus, the value 7,000 contains only one significant figure. However, adding a decimal point and another zero gives the number 7,000.0, which has five significant figures.

A good way to measure the significance of zeros, if the above rules become confusing, is to convert the number to the exponential form. If the zeros can be omitted, then they are not significant. For example, 7000 expressed in exponential form is 7×10^3 and contains one significant figure. With 7000.0, the zeros are retained and the number becomes 7.0000×10^3. If we were to convert 0.007 to exponential form, the value is 7×10^{-3} and only one significant figure is indicated. As a rule, determining significant figures in arithmetic operations is dictated by the value having the least number of significant figures. The easiest way to avoid any confusion is to perform all the calculations and then round off the final answer to the appropriate digits. For example, $36.54 \times 238 \times 1.1 = 9566.172$, and because 1.1 contains only two significant figures, the answer would be reported as 9600 (remember, the two zeros are not significant). This method works fine for most calculations, except when adding or subtracting numbers containing decimals. In those cases, the number of significant figures in the final value is determined by the numbers that follow the decimal point. Thus, when adding $7.45 + 8.725 = 16.175$, the sum is rounded to 16.18 because 7.45 has only two numbers after the decimal point. Likewise, 433.8–32.66 gives 401.14, which rounds off to 401.1.

A word of caution is warranted when using the simple rule stated above, for there is a tendency to underestimate the significant figures in the final answer. For example, take the situation in which we determined the caffeine in an unknown solution to be 43.5 ppm (see Eq. 4.24). We had to dilute the sample 50-fold using a volumetric flask in order to fit the unknown within the range of our method. To calculate the caffeine in the original sample, we multiply our result by 50 or $43.5 \times 50 = 2{,}175\ \mu g/mL$ in the unknown. Based on our rule above, we then would round off the number to one significant figure (because 50 contains one significant figure) and report the value as 2,000. However, doing this actually underestimates the sensitivity of our procedure, because we ignore the accuracy of the volumetric flask used for the dilution. A Class-A volumetric flask has a tolerance of 0.05 mL; thus, a more reasonable way to express the dilution factor would be 50.0 instead of 50. We now have increased the significant figures in the answer by two, and the value becomes $2{,}180\ \mu/mL$.

As you can see, an awareness of significant figures and how they are adopted requires close inspection. The guidelines can be helpful, but they do not always work unless each individual value or number is closely inspected.

4.5.2 Outlier Data and Testing [2, 3]

Inevitably, during the course of working with experimental data, we will come across outlier values that do not match the others. Can you reject that value and thus not use it in calculating the final reported results?

The answer is "very rarely" and only after careful consideration. If you are routinely rejecting data to help make your assay look better, then you are misrepresenting the results and the precision of the assay. If the bad value resulted from an identifiable mistake in that particular test, then it is probably safe to drop the value. Again, caution is advised because you may be rejecting a value that is closer to the true value than some of the other values.

Consistently poor accuracy or precision indicates that an improper technique or incorrect reagent was used or that the test was not very good. It is best to make changes in the procedure or change methods rather than try to figure out ways to eliminate undesirable values.

There are several tests for rejecting outlier data. In addition, the use of more robust statistical estimators of the population mean can minimize effects of extreme outlier values [2]. The simplest test for rejecting outlier data is the Dixon Q test [7, 8], often called just the Q-test. The advantage is that this test can be easily calculated with a simple calculator and is useful for a small group of data. In this test, a **Q-value** is calculated as shown below and compared to values in a table. If the calculated value is larger than the table value, then the questionable measurement can be rejected at the 90 % confidence level:

$$Q-\text{value} = x_2 - x_1 / W \qquad (4.26)$$

where:

x_1 = the questionable value

4.4

table | *Q*-values for the rejection of results

Number of observations	Q of rejection (90% level)
3	0.94
4	0.76
5	0.64
6	0.56
7	0.51
8	0.47
9	0.44
10	0.41

Adapted from Dean and Dixon [7]

x_2 = the next closest value to x_1
W = the total spread of all values, obtained by subtracting the lowest value from the highest value

Table 4.4 provides the rejection of *Q*-values for a 90% confidence level.

The example below shows how the test is used for the moisture level of uncooked hamburger for which four replicates were performed giving values of 64.53, 64.45, 64.78, and 55.31. The 55.31 value looks as if it is too low compared to the other results. Can that value be rejected? For our example, x_1 is the questionable value (55.31) and x_2 is the closest neighbor to x_1 (which is 64.45). The spread (*W*) is the high value minus the low measurement, which is 64.78–55.31:

$$Q-\text{value} = \frac{64.45 - 55.31}{64.78 - 55.31} = \frac{9.14}{9.47} = 0.97 \qquad (4.27)$$

From Table 4.4, we see that the calculated *Q*-value must be greater than 0.76 to reject the data. Thus, we make the decision to reject the 55.31% moisture value and do not use it in calculating the mean.

4.6 SUMMARY

This chapter focuses on statistical methods to measure data variability, precision, etc. and basic mathematical treatment that can be used in evaluating a group of data. For example, it should be almost second nature to determine a mean, standard deviation, and CV when evaluating replicate analyses of an individual sample. In evaluating linear standard curves, best line fits should always be determined along with the indicators of the degree of linearity (correlation coefficient or coefficient of determination). Fortunately, most computer spreadsheet and graphics software will readily perform the calculations for you. Guidelines are available to enable one to report analytical results in a way that tells something about the sensitivity and confidence of a particular test. A section is included which describes sensitivity and limit of detection as related to various analytical methods and regulatory agency policies. Additional information includes the proper use of significant figures, rules for rounding off numbers, and use of various tests to reject grossly aberrant individual values (outliers).

4.7 STUDY QUESTIONS

1. Method A to quantitate a particular food component was reported to be more specific and accurate than Method B, but Method A had lower precision. Explain what this means.
2. You are considering adopting a new analytical method in your lab to measure moisture content of cereal products. How would you determine the precision of the new method and compare it to the old method? Include any equations to be used for any needed calculations.
3. A sample known to contain 20 g/L glucose is analyzed by two methods. Ten determinations were made for each method and the following results were obtained:

Method A	Method B
Mean = 19.6	Mean = 20.2
Std. Dev. = 0.055	Std. Dev. = 0.134

   (a) Precision and accuracy:

      (i) Which method is more precise? Why do you say this?
      (ii) Which method is more accurate? Why do you say this?

   (b) In the equation to determine the standard deviation, $n-1$ was used rather than just n. Would the standard deviation have been smaller or larger for each of those values above if simply n had been used?

   (c) You have determined that values obtained using Method B should not be accepted if outside the range of two standard deviations from the mean. What range of values will be acceptable?

   (d) Do the data above tell you anything about the specificity of the method? Describe what "specificity" of the method means as you explain your answer.

4. Differentiate "standard deviation" from "coefficient of variation," "standard error of the mean," and "confidence interval."
5. Differentiate the terms "absolute error" versus "relative error." Which is more useful? Why?

6. For each of the errors described below in performing an analytical procedure, classify the error as random error, systematic error, or blunder, and describe a way to overcome the error:

   (a) Automatic pipettor consistently delivered 0.96 mL rather than 1.00 mL.
   (b) Substrate was not added to one tube in an enzyme assay.

7. Differentiate the terms "sensitivity" and "limit of detection."

8. The correlation coefficient for standard curve A is reported as 0.9970. The coefficient of determination for standard curve B is reported as 0.9950. In which case do the data better fit a straight line?

4.8 PRACTICE PROBLEMS

1. How many significant figures are in the following numbers: 0.0025, 4.50, 5.607?

2. What is the correct answer for the following calculation expressed in the proper amount of significant figures?

$$\frac{2.43 \times 0.01672}{1.83215} =$$

3. Given the following data on dry matter (88.62, 88.74, 89.20, 82.20), determine the mean, standard deviation, and CV. Is the precision for this set of data acceptable? Can you reject the value 82.20 since it seems to be different than the others? What is the 95% confidence level you would expect your values to fall within if the test were repeated? If the true value for dry matter is 89.40, what is the percent relative error?

4. Compare the two groups of standard curve data below for sodium determination by atomic emission spectroscopy. Draw the standard curves using graph paper or a computer software program. Which group of data provides a better standard curve? Note that the absorbance of the emitted radiation at 589 nm increases proportionally to sodium concentration. Calculate the amount of sodium in a sample with a value of 0.555 for emission at 589 nm. Use both standard curve groups and compare the results.

Sodium concentration (µg/mL)	Emission at 589 nm
Group A—sodium standard curve	
1.00	0.050
3.00	0.140
5.00	0.242
10.0	0.521
20.0	0.998

Sodium concentration (µg/mL)	Emission at 589 nm
Group B—sodium standard curve	
1.00	0.060
3.00	0.113
5.00	0.221
10.00	0.592
20.00	0.917

Answers

1. 2, 3, 4
2. 0.0222
3. Mean = 87.19, SD_{n-1} = 3.34:

$$CV = \frac{3.34}{87.18} \times 100\% = 3.83\%$$

Thus, the precision is acceptable because it is less than 5%:

$$Q - \text{calc value} = \frac{88.62 - 82.20}{89.20 - 82.20} = \frac{6.42}{7.00} = 0.917$$

Q_{calc} = 0.92; therefore, the value 82.20 can be rejected because it is more than 0.76 from Table 4.4, using 4 as number of observations:

$$CI\,(at\,95\%) = 87.19 \pm 3.18 \times \frac{3.34}{\sqrt{4}}$$
$$= 87.19 \pm 5.31$$

Relative error = $\%E_{rel}$ where mean is 87.19 and true value is 89.40:

$$\%E_{rel} = \frac{\bar{x} - T}{T} \times 100\% = \frac{87.19 - 89.40}{89.40} \times 100\%$$
$$= -2.47\%$$

4. Using linear regression, we get

   Group A: $y = 0.0504x - 0.0029$, $r^2 = 0.9990$.
   Group B: $y = 0.0473x + 0.0115$, $r^2 = 0.9708$.
   Group A r^2 is closer to 1.000 and is more linear and thus the better standard curve.

   Sodium in the sample using group A standard curve is

   $$0.555 = 0.0504x - 0.0029, \quad x = 11.1\mu g\,/\,mL$$

   Sodium in the sample using group B standard curve is

   $$0.555 = 0.0473x + 0.0115, \quad x = 11.5\mu g\,/\,mL$$

REFERENCES

1. Garfield F M, Klestra E, and Hirsch J (2000) Quality assurance principles for analytical laboratories, AOAC International, Gaithersburg, MD. (This book covers mostly quality assurance issues, but Chap. 3 is a good chapter on sampling and on the basics of statistical applications to data.)

2. Ellison, SLR, Barwick, VJ, Farrant, TJD (2009) Practical Statistics for the Analytical Scientist: A Bench Guide, Chapter 5, 2nd ed. RSC Publishing, Cambridge, UK. (This an excellent text for beginner and advanced analytical chemists alike. It contains a fair amount of detail, sufficient for most analytical statistics, yet works through the material starting at a basic introductory level. The authors have an entire chapter, Chap. 5, devoted to the various test used to evaluating and rejecting data outliers.)

3. Miller JN, Miller JC (2010) Statistics and chemometrics for analytical chemistry, 6th edn. Pearson Prentice Hall, Upper Saddle River, NJ. (This is another excellent introductory text for beginner analytical chemists at the undergraduate and graduate level. It contains a fair amount of detail, sufficient for most analytical statistics, yet works through the material starting at a basic introductory level. Chapter 1 covers the types of experimental error. The assay evaluation parameters LOD, MDL, and LOQ are thoroughly discussed. The authors also cover outlier testing including the Q-test used for rejecting data.)

4. Skoog DA, West DM, Holler JF, Crouch SR (2014) Fundamentals of analytical chemistry, 9th edn. Cengage Brain, Independence, KY. (Part I, Chaps. 5–8, have thorough coverage of the statistics needed by an analytical chemist in an easy-to-read style.)

5. Shrivastava A, Gupta VB (2011) Methods for the determination of limit of detection and limit of quantitation of the analytical methods. Chron Young Sci 2(1):21–25

6. Code of Federal Regulations (2011) Environmental Protection Agency. 40 CFR, Vol. 22. Chapter I part 136 – Guidelines Establishing Test Procedures for the Analysis of Pollutants. Appendix B to Part 136-Definition and Procedure for the Determination of the Method Detection Limit-Revision 1.11

7. Dean RB, Dixon WJ (1951) Simplified statistics for small numbers of observations. Anal Chem 23(4): 636–638

8. Rorabacher DB (1991) Statistical treatment for rejection of deviant values: critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence level. Anal Chem 63 (2):139–146