

Chapter 14

Developing a Standard Method

Chapter Overview

- 14A Optimizing the Experimental Procedure
- 14B Verifying the Method
- 14C Validating the Method as a Standard Method
- 14D Using Excel and R for an Analysis of Variance
- 14E Key Terms
- 14F Chapter Summary
- 14G Problems
- 14H Solutions to Practice Exercises

In Chapter 1 we made a distinction between analytical chemistry and chemical analysis. Among the goals of analytical chemistry are improving established methods of analysis, extending existing methods of analysis to new types of samples, and developing new analytical methods. Once we develop a new method, its routine application is best described as chemical analysis. We recognize the status of these established methods by calling them standard methods.

Numerous examples of standard methods are presented and discussed in Chapters 8–13. What we have yet to consider is what constitutes a standard method. In this chapter we discuss how we develop a standard method, including optimizing the experimental procedure, verifying that the method produces acceptable precision and accuracy in the hands of a single analyst, and validating the method for general use.

14A Optimizing the Experimental Procedure

In the presence of H_2O_2 and H_2SO_4 , a solution of vanadium forms a reddish brown color that is believed to be a compound with the general formula $(\text{VO})_2(\text{SO}_4)_3$. The intensity of the solution's color depends on the concentration of vanadium, which means we can use its absorbance at a wavelength of 450 nm to develop a quantitative method for vanadium.

The intensity of the solution's color also depends on the amounts of H_2O_2 and H_2SO_4 that we add to the sample—in particular, a large excess of H_2O_2 decreases the solution's absorbance as it changes from a reddish brown color to a yellowish color.¹ Developing a **STANDARD METHOD** for vanadium based on this reaction requires that we optimize the amount of H_2O_2 and H_2SO_4 added to maximize the absorbance at 450 nm. Using the terminology of statisticians, we call the solution's absorbance the system's **RESPONSE**. Hydrogen peroxide and sulfuric acid are **FACTORS** whose concentrations, or **FACTOR LEVELS**, determine the system's response. To optimize the method we need to find the best combination of factor levels. Usually we seek a maximum response, as is the case for the quantitative analysis of vanadium as $(\text{VO})_2(\text{SO}_4)_3$. In other situations, such as minimizing an analysis's percent error, we seek a minimum response.

We will return to this analytical method for vanadium in [Example 14.4](#) and in [Problem 14.11](#).

14A.1 Response Surfaces

One of the most effective ways to think about an optimization is to visualize how a system's response changes when we increase or decrease the levels of one or more of its factors. We call a plot of the system's response as a function of the factor levels a **RESPONSE SURFACE**. The simplest response surface has one factor and is drawn in two dimensions by placing the responses on the y -axis and the factor's levels on the x -axis. The calibration curve in Figure 14.1 is an example of a one-factor response surface. We also can define the response surface mathematically. The response surface in Figure 14.1, for example, is

$$A = 0.008 + 0.0896C_A$$

where A is the absorbance and C_A is the analyte's concentration in ppm.

For a two-factor system, such as the quantitative analysis for vanadium described earlier, the response surface is a flat or curved plane in three dimensions. As shown in [Figure 14.2a](#), we place the response on the z -axis and the factor levels on the x -axis and the y -axis. [Figure 14.2a](#) shows a pseudo-three dimensional wireframe plot for a system that obeys the equation

$$R = 3.0 - 0.30A + 0.020AB$$

where R is the response, and A and B are the factors. We also can represent a two-factor response surface using the two-dimensional level plot in [Figure 12.4b](#), which uses a color gradient to show the response on a two-

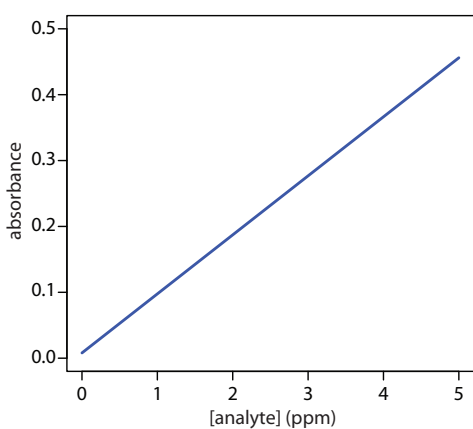


Figure 14.1 A calibration curve is an example of a one-factor response surface. The responses (absorbance) are plotted on the y -axis and the factor levels (concentration of analyte) are plotted on the x -axis.

Another name for a level plot is a heat-map.

¹ *Vogel's Textbook of Quantitative Inorganic Analysis*, Longman: London, 1978, p. 752.

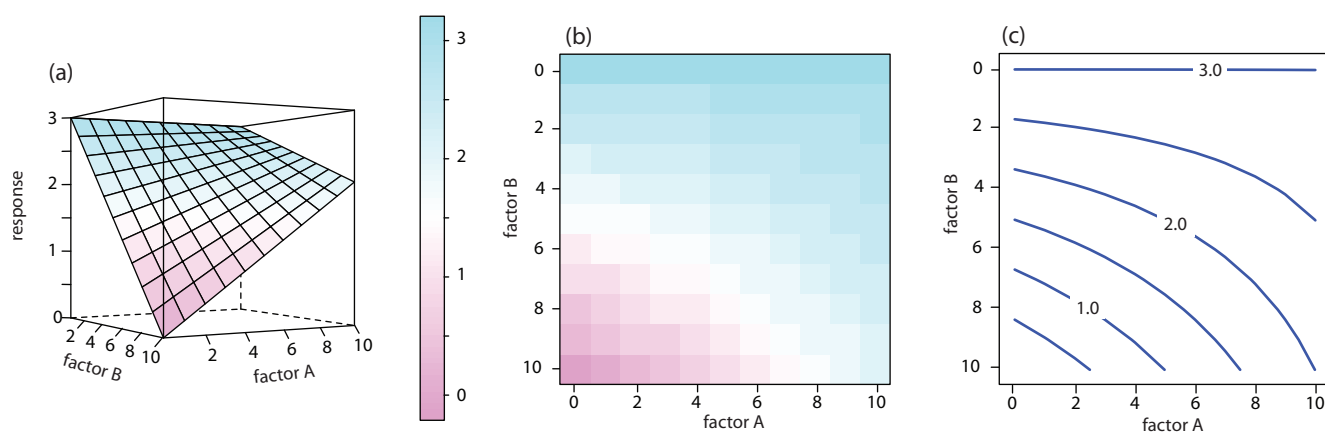


Figure 14.2 Three examples of a two-factor response surface displayed as (a) a pseudo-three-dimensional wireframe plot, (b) a two-dimensional level plot, and (c) a two-dimensional contour plot. We call the display in (a) a pseudo-three dimensional response surface because we show the presence of three dimensions on the page's flat, two-dimensional surface.

dimensional grid, or using the two-dimensional contour plot in Figure 14.2c, which uses contour lines to display the response surface.

The response surfaces in Figure 14.2 cover a limited range of factor levels ($0 \leq A \leq 10$, $0 \leq B \leq 10$), but we can extend each to more positive or to more negative values because there are no constraints on the factors. Most response surfaces of interest to an analytical chemist have natural constraints imposed by the factors, or have practical limits set by the analyst. The response surface in Figure 14.1, for example, has a natural constraint on its factor because the analyte's concentration cannot be less than zero.

If we have an equation for the response surface, then it is relatively easy to find the optimum response. Unfortunately, when developing a new analytical method, we rarely know any useful details about the response surface. Instead, we must determine the response surface's shape and locate its optimum response by running appropriate experiments. The focus of this section is on useful experimental methods for characterizing a response surface. These experimental methods are divided into two broad categories: searching methods, in which an algorithm guides a systematic search for the optimum response, and modeling methods, in which we use a theoretical model or an empirical model of the response surface to predict the optimum response.

14A.2 Searching Algorithms for Response Surfaces

Figure 14.3 shows a portion of the South Dakota Badlands, a barren landscape that includes many narrow ridges formed through erosion. Suppose you wish to climb to the highest point on this ridge. Because the shortest path to the summit is not obvious, you might adopt the following simple rule: look around you and take one step in the direction that has the greatest change in elevation, and then repeat until no further step is possible. The route you follow is the result of a systematic search that uses a **SEARCHING ALGORITHM**. Of course there are as many possible routes as there are starting

We also can overlay a level plot and a contour plot. See Figure 14.7b for a typical example.

We express this constraint as $C_A \geq 0$.

Searching algorithms have names: the one described here is the method of steepest ascent.

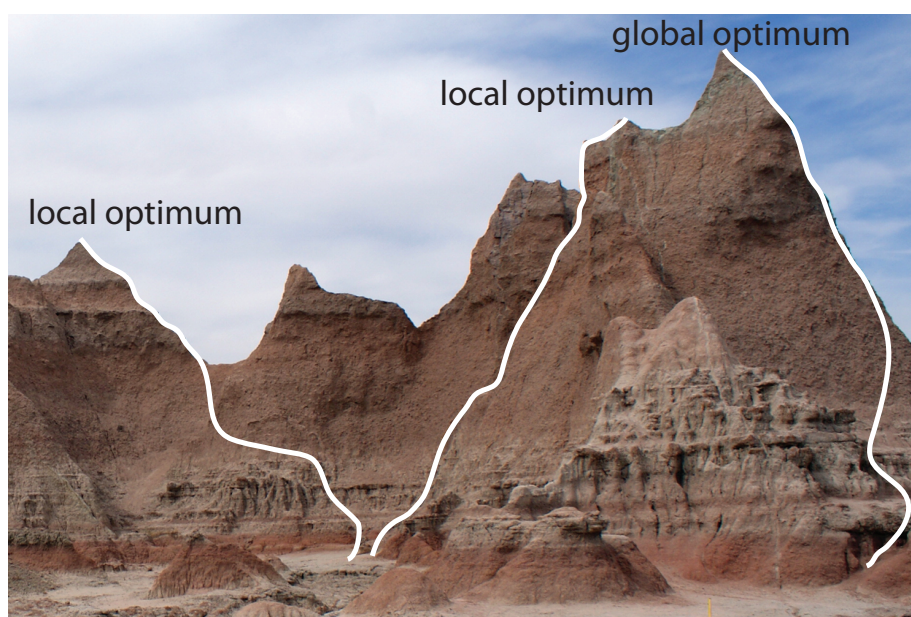


Figure 14.3 Finding the highest point on a ridge using a searching algorithm is one useful method for finding the optimum on a response surface. The path on the far right reaches the highest point, or the global optimum. The other two paths reach local optima. This ridge is part of the South Dakota Badlands National Park. You can read about the geology of the park at www.nps.gov/badl/.

points, three examples of which are shown in Figure 14.3. Note that some routes do not reach the highest point—what we call the **GLOBAL OPTIMUM**. Instead, many routes reach a **LOCAL OPTIMUM** from which further movement is impossible.

We can use a systematic searching algorithm to locate the optimum response for an analytical method. We begin by selecting an initial set of factor levels and measure the response. Next, we apply the rules of our searching algorithm to determine a new set of factor levels and measure its response, continuing this process until we reach an optimum response. Before we consider two common searching algorithms, let's consider how we evaluate a searching algorithm.

EFFECTIVENESS AND EFFICIENCY

A searching algorithm is characterized by its effectiveness and its efficiency. To be **EFFECTIVE**, a searching algorithm must find the response surface's global optimum, or at least reach a point near the global optimum. A searching algorithm may fail to find the global optimum for several reasons, including a poorly designed algorithm, uncertainty in measuring the response, and the presence of local optima. Let's consider each of these potential problems.

A poorly designed algorithm may prematurely end the search before it reaches the response surface's global optimum. As shown in Figure 14.4, when climbing a ridge that slopes up to the northeast, an algorithm is likely

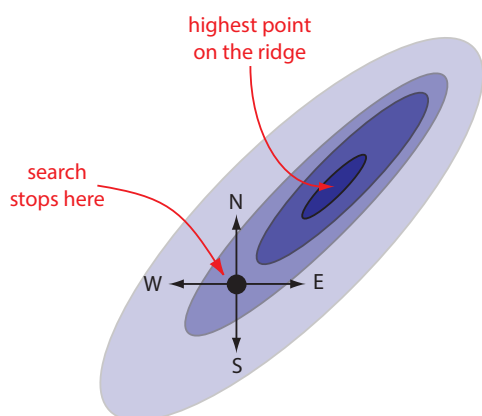


Figure 14.4 Example showing how a poorly designed searching algorithm can fail to find a response surface's global optimum.

to fail it if limits your steps only to the north, south, east, or west. An algorithm that cannot respond to a change in the direction of steepest ascent is not an effective algorithm.

All measurements contain uncertainty, or noise, that affects our ability to characterize the underlying signal. When the noise is greater than the local change in the signal, then a searching algorithm is likely to end before it reaches the global optimum. Figure 14.5 provides a different view of Figure 14.3, which shows us that the relatively flat terrain leading up to the ridge is heavily weathered and very uneven. Because the variation in local height (the noise) exceeds the slope (the signal), our searching algorithm ends the first time we step up onto a less weathered local surface.

Finally, a response surface may contain several local optima, only one of which is the global optimum. If we begin the search near a local optimum, our searching algorithm may never reach the global optimum. The ridge in Figure 14.3, for example, has many peaks. Only those searches that begin at the far right will reach the highest point on the ridge. Ideally, a searching algorithm should reach the global optimum regardless of where it starts.

A searching algorithm always reaches an optimum. Our problem, of course, is that we do not know if it is the global optimum. One method for evaluating a searching algorithm's effectiveness is to use several sets of initial factor levels, find the optimum response for each, and compare the results. If we arrive at or near the same optimum response after starting from very different locations on the response surface, then we are more confident that it is the global optimum.

EFFICIENCY is a searching algorithm's second desirable characteristic. An efficient algorithm moves from the initial set of factor levels to the optimum response in as few steps as possible. In seeking the highest point on the ridge in Figure 14.5, we can increase the rate at which we approach the optimum by taking larger steps. If the step size is too large, however, the difference between the experimental optimum and the true optimum may be unacceptably large. One solution is to adjust the step size during the search, using larger steps at the beginning and smaller steps as we approach the global optimum.

ONE-FACTOR-AT-A-TIME OPTIMIZATION

A simple algorithm for optimizing the quantitative method for vanadium described earlier is to select initial concentrations for H_2O_2 and H_2SO_4 and measure the absorbance. Next, we optimize one reagent by increasing or decreasing its concentration—holding constant the second reagent's concentration—until the absorbance decreases. We then vary the concentration of the second reagent—maintaining the first reagent's optimum concentration—until we no longer see an increase in the absorbance. We can stop this process, which we call a **ONE-FACTOR-AT-A-TIME OPTIMIZATION**, after one cycle or repeat the steps until the absorbance reaches a maximum value or it exceeds an acceptable threshold value.



Figure 14.5 Another view of the ridge in Figure 14.3 that shows the weathered terrain leading up to the ridge. The yellow rod at the bottom of the figure, which marks the trail, is about 18 in high.

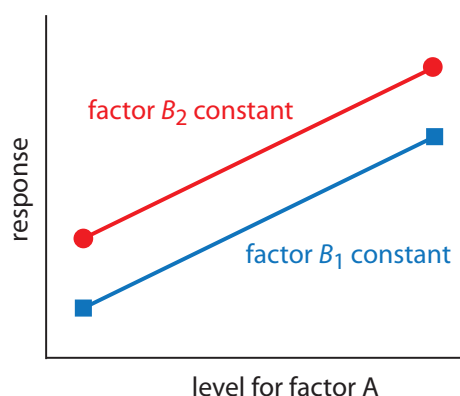


Figure 14.6 Factor effect plot for two independent factors. Note that the two lines are parallel, indicating that the level for factor B does not influence how factor A 's level affects the response.

Practice Exercise 14.1

Using the data in Table 14.1, show that factor B 's affect on the response is independent of factor A .

Click [here](#) to review your answer to this exercise.

Table 14.1 Example of Two Independent Factors

factor A	factor B	response
A_1	B_1	40
A_2	B_1	80
A_1	B_2	60
A_2	B_2	100

A one-factor-at-a-time optimization is consistent with a notion that to determine the influence of one factor we must hold constant all other factors. This is an effective, although not necessarily an efficient experimental design when the factors are independent.² Two factors are **INDEPENDENT** when a change in the level of one factor does not influence the effect of a change in the other factor's level. Table 14.1 provides an example of two independent factors. If we hold factor B at level B_1 , changing factor A from level A_1 to level A_2 increases the response from 40 to 80, or a change in response, ΔR , of

$$R = 80 - 40 = 40$$

If we hold factor B at level B_2 , we find that we have the same change in response when the level of factor A changes from A_1 to A_2 .

$$R = 100 - 60 = 40$$

We can see this independence visually if we plot the response as a function of factor A 's level, as shown in Figure 14.6. The parallel lines show that the level of factor B does not influence factor A 's effect on the response.

Mathematically, two factors are independent if they do not appear in the same term in the equation that describes the response surface. Equation 14.1, for example, describes a response surface with independent factors because no term in the equation includes both factor A and factor B .

$$R = 2.0 + 0.12A + 0.48B - 0.03A^2 - 0.03B^2 \quad 14.1$$

Figure 14.7 shows the resulting pseudo-three-dimensional surface and a contour map for equation 14.1.

The easiest way to follow the progress of a searching algorithm is to map its path on a contour plot of the response surface. Positions on the response surface are identified as (a, b) where a and b are the levels for factor A and for factor B . The contour plot in Figure 14.7b, for example, shows four one-factor-at-a-time optimizations of the response surface for equation 14.1. The effectiveness and efficiency of this algorithm when optimizing independent factors is clear—each trial reaches the optimum response at $(2, 8)$ in a single cycle.

Unfortunately, factors often are not independent. Consider, for example, the data in Table 14.2 where a change in the level of factor B from

² Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*, Wiley-Interscience: New York, 1986.

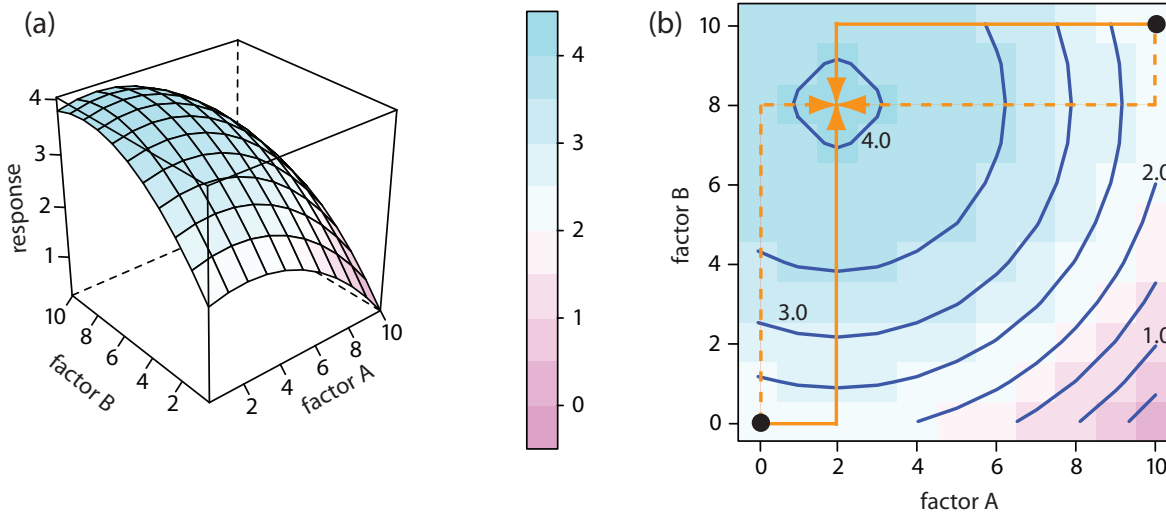


Figure 14.7 The response surface for two independent factors based on equation 14.1, displayed as (a) a wireframe, and as (b) an overlaid contour plot and level plot. The orange lines in (b) show the progress of one-factor-at-a-time optimizations beginning from two starting points (•) and optimizing factor *A* first (solid line) or factor *B* first (dashed line). All four trials reach the optimum response of (2,8) in a single cycle.

level B_1 to level B_2 has a significant effect on the response when factor *A* is at level A_1

$$R = 60 - 20 = 40$$

but no effect when factor *A* is at level A_2 .

$$R = 80 - 80 = 0$$

Figure 14.8 shows this **DEPENDENT** relationship between the two factors. Factors that are dependent are said to interact and the equation for the response surface' includes an interaction term that contains both factor *A* and factor *B*. The final term in equation 14.2, for example, accounts for the interaction between factor *A* and factor *B*.

$$R = 5.5 + 1.5A + 0.6B - 0.15A^2 - 0.0245B^2 - 0.0857AB \tag{14.2}$$

Figure 14.9 shows the resulting pseudo-three-dimensional surface and a contour map for equation 14.2.

The progress of a one-factor-at-a-time optimization for equation 14.2 is shown in Figure 14.9b. Although the optimization for dependent factors is effective, it is less efficient than that for independent factors. In this case

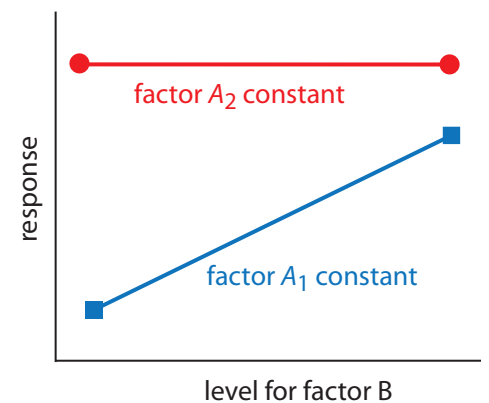


Figure 14.8 Factor effect plot for two dependent factors. Note that the two lines are not parallel, indicating that the level for factor *A* influences how factor *B*'s level affects the response.

Table 14.2 Example of Two Dependent Factors		
factor A	factor B	response
A_1	B_1	20
A_2	B_1	80
A_1	B_2	60
A_2	B_2	80

Practice Exercise 14.2

Using the data in Table 14.2, show that factor *A*'s affect on the response is independent of factor *B*.

Click [here](#) to review your answer to this exercise.

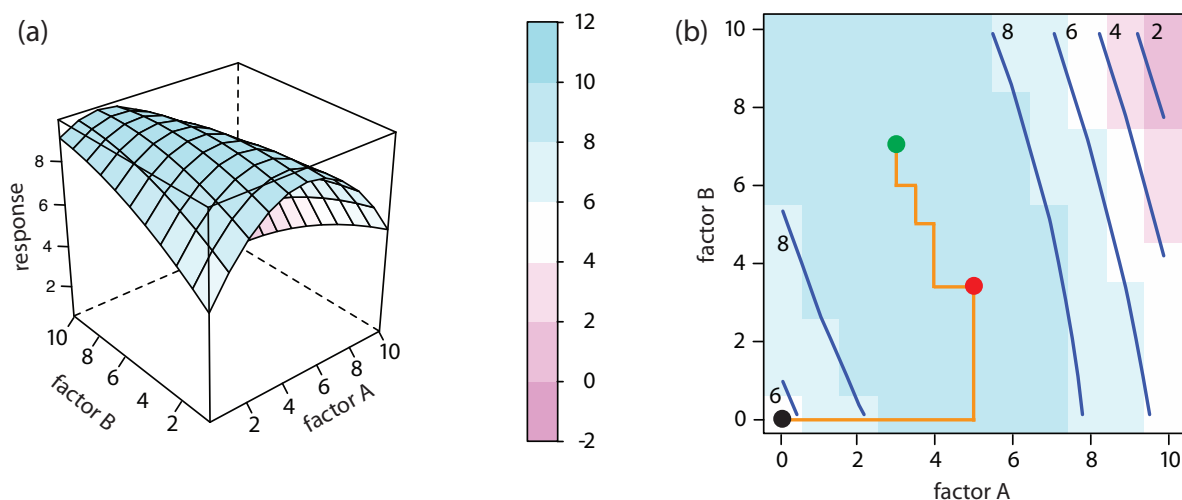


Figure 14.9 The response surface for two dependent factors based on [equation 14.2](#), displayed as (a) a wireframe, and as (b) an overlaid contour plot and level plot. The **orange** lines in (b) show the progress of one-factor-at-a-time optimization beginning from the starting point (•) and optimizing factor *A* first. The red dot (•) marks the end of the first cycle. It takes four cycles to reach the optimum response of (3, 7) as shown by the green dot (•).

it takes four cycles to reach the optimum response of (3, 7) if we begin at (0, 0).

SIMPLEX OPTIMIZATION

One strategy for improving the efficiency of a searching algorithm is to change more than one factor at a time. A convenient way to accomplish this when there are two factors is to begin with three sets of initial factor levels as the vertices of a triangle. After measuring the response for each set of factor levels, we identify the combination that gives the worst response and replace it with a new set of factor levels using a set of rules (Figure 14.10). This process continues until we reach the global optimum or until no further optimization is possible. The set of factor levels is called a **SIMPLEX**. In general, for k factors a simplex is a $k + 1$ dimensional geometric figure.³

Thus, for two factors the simplex is a triangle. For three factors the simplex is a tetrahedron.

³ (a) Spendley, W.; Hext, G. R.; Himsworth, F. R. *Technometrics* **1962**, *4*, 441–461; (b) Deming, S. N.; Parker, L. R. *CRC Crit. Rev. Anal. Chem.* **1978** *7*(3), 187–202.

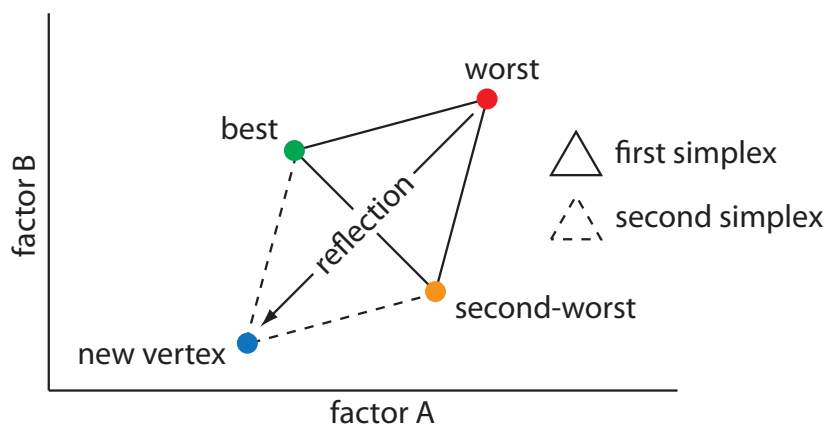


Figure 14.10 Example of a two-factor simplex. The original simplex is formed by the **green**, **orange**, and **red** vertices. Replacing the worst vertex with a new vertex moves the simplex to a new position on the response surface.

To place the initial two-factor simplex on the response surface, we choose a starting point (a, b) for the first vertex and place the remaining two vertices at $(a + s_a, b)$ and $(a + 0.5s_a, b + 0.87s_b)$ where s_a and s_b are step sizes for factor A and for factor B .⁴ The following set of rules moves the simplex across the response surface in search of the optimum response:

Rule 1. Rank the vertices from best (v_b) to worst (v_w).

Rule 2. Reject the worst vertex (v_w) and replace it with a new vertex (v_n) by reflecting the worst vertex through the midpoint of the remaining vertices. The new vertex's factor levels are twice the average factor levels for the retained vertices minus the factor levels for the worst vertex. For a two-factor optimization, the equations are shown here where v_s is the third vertex.

$$a_{v_n} = 2\left(\frac{a_{v_b} + a_{v_s}}{2}\right) - a_{v_w} \quad 14.3$$

$$b_{v_n} = 2\left(\frac{b_{v_b} + b_{v_s}}{2}\right) - b_{v_w} \quad 14.4$$

The variables a and b in equation 14.3 and equation 14.4 are the factor levels for factor A and for factor B , respectively. [Problem 14.3](#) in the end-of-chapter problems asks you to derive these equations.

Rule 3. If the new vertex has the worst response, then return to the previous vertex and reject the vertex with the second worst response, (v_s) calculating the new vertex's factor levels using rule 2. This rule ensures that the simplex does not return to the previous simplex.

Rule 4. Boundary conditions are a useful way to limit the range of possible factor levels. For example, it may be necessary to limit a factor's concentration for solubility reasons, or to limit the temperature because a reagent is thermally unstable. If the new vertex exceeds a boundary condition, then assign it the worst response and follow rule 3.

Because the size of the simplex remains constant during the search, this algorithm is called a **FIXED-SIZED SIMPLEX OPTIMIZATION**. Example 14.1 illustrates the application of these rules.

Example 14.1

Find the optimum for the response surface in [Figure 14.9](#) using the fixed-sized simplex searching algorithm. Use $(0, 0)$ for the initial factor levels and set each factor's step size to 1.00.

SOLUTION

Letting $a = 0$, $b = 0$, $s_a = 1.00$, and $s_b = 1.00$ gives the vertices for the initial simplex as

$$\text{vertex 1: } (a, b) = (0, 0)$$

$$\text{vertex 2: } (a + s_a, b) = (1.00, 0)$$

$$\text{vertex 3: } (a + 0.5s_a, b + 0.87s_b) = (0.50, 0.87)$$

⁴ Long, D. E. *Anal. Chim. Acta* **1969**, *46*, 193–206.

The size of the initial simplex ultimately limits the effectiveness and the efficiency of a fixed-size simplex searching algorithm. We can increase its efficiency by allowing the size of the simplex to expand or to contract in response to the rate at which we approach the optimum. For example, if we find that a new vertex is better than any of the vertices in the preceding simplex, then we expand the simplex further in this direction on the assumption that we are moving directly toward the optimum. Other conditions might cause us to contract the simplex—to make it smaller—to encourage the optimization to move in a different direction. We call this a **VARIABLE-SIZED SIMPLEX OPTIMIZATION**.

Consult this chapter's additional resources for further details of the variable-sized simplex optimization.

The responses, from [equation 14.2](#), for the three vertices are shown in the following table

vertex	a	b	response
v_1	0	0	5.50
v_2	1.00	0	6.85
v_3	0.50	0.87	6.68

with v_1 giving the worst response and v_3 the best response. Following Rule 1, we reject v_1 and replace it with a new vertex using [equation 14.3](#) and [equation 14.4](#); thus

$$a_{v_4} = 2 \times \frac{1.00 + 0.50}{2} - 0 = 1.50$$

$$b_{v_4} = 2 \times \frac{0 + 0.87}{2} - 0 = 0.87$$

The following table gives the vertices of the second simplex.

vertex	a	b	response
v_2	1.50	0	6.85
v_3	0.50	0.87	6.68
v_4	1.50	0.87	7.80

with v_3 giving the worst response and v_4 the best response. Following Rule 1, we reject v_3 and replace it with a new vertex using [equation 14.3](#) and [equation 14.4](#); thus

$$a_{v_5} = 2 \times \frac{1.00 + 1.50}{2} - 0.50 = 2.00$$

$$b_{v_5} = 2 \times \frac{0 + 0.87}{2} - 0.87 = 0$$

The following table gives the vertices of the third simplex.

vertex	a	b	response
v_2	1.50	0	6.85
v_4	1.50	0.87	7.80
v_5	2.00	0	7.90

The calculation of the remaining vertices is left as an exercise. [Figure 14.11](#) shows the progress of the complete optimization. After 29 steps the simplex begins to repeat itself, circling around the optimum response of (3, 7).

14A.3 Mathematical Models of Response Surfaces

A response surface is described mathematically by an equation that relates the response to its factors. [Equation 14.1](#) and [equation 14.2](#) provide two examples of such mathematical models. If we measure the response for several combinations of factor levels, then we can model the response surface by

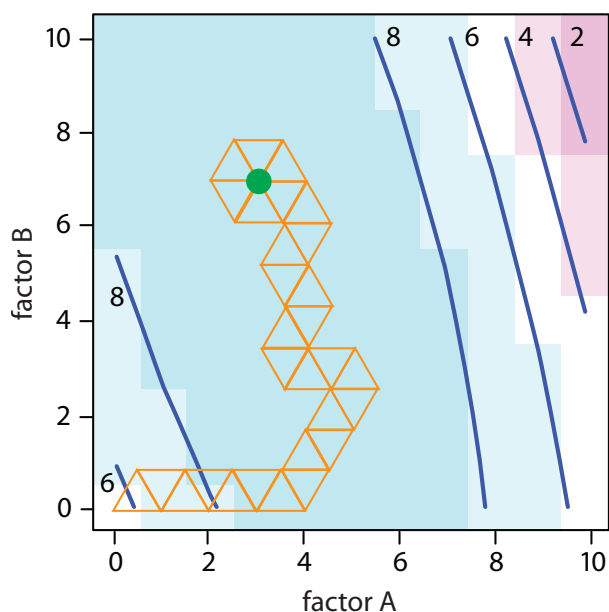


Figure 14.11 Progress of the fixed-size simplex optimization in [Example 14.1](#). The green dot (•) marks the optimum response of (3, 7). Optimization ends when the simplexes begin to circle around a single vertex.

using a regression analysis to fit an appropriate equation to the data. There are two broad categories of models that we can use for a regression analysis: theoretical models and empirical models.

THEORETICAL MODELS OF THE RESPONSE SURFACE

A **THEORETICAL MODEL** is derived from the known chemical and physical relationships between the response and its factors. In spectrophotometry, for example, Beer's law is a theoretical model that relates an analyte's absorbance, A , to its concentration, C_A

$$A = \varepsilon b C_A$$

where ε is the molar absorptivity and b is the pathlength of the electromagnetic radiation passing through the sample. A Beer's law calibration curve, therefore, is a theoretical model of a response surface.

EMPIRICAL MODELS OF THE RESPONSE SURFACE

In many cases the underlying theoretical relationship between the response and its factors is unknown. We still can develop a model of the response surface if we make some reasonable assumptions about the underlying relationship between the factors and the response. For example, if we believe that the factors A and B are independent and that each has only a first-order effect on the response, then the following equation is a suitable model.

$$R = \beta_0 + \beta_a A + \beta_b B$$

where R is the response, A and B are the factor levels, and β_0 , β_a , and β_b are adjustable parameters whose values are determined by a linear regression analysis. Other examples of equations include those for dependent factors

$$R = \beta_0 + \beta_a A + \beta_b B + \beta_{ab} AB$$

For a review of Beer's law, see Section 10B.3 in Chapter 10. [Figure 14.1](#) is an example of a Beer's law calibration curve.

The calculations for a linear regression when the model is first-order in one factor (a straight line) are described in Chapter 5D. A complete mathematical treatment of linear regression for models that are second-order in one factor or which contain more than one factor is beyond the scope of this text. The computations for a few special cases, however, are straightforward and are considered in this section. A more comprehensive treatment of linear regression is available in several of this chapter's additional resources.

and those with higher-order terms.

$$R = \beta_0 + \beta_a A + \beta_b B + \beta_{aa} A^2 + \beta_{bb} B^2$$

Each of these equations provides an **EMPIRICAL MODEL** of the response surface because it has no basis in a theoretical understanding of the relationship between the response and its factors. Although an empirical model may provide an excellent description of the response surface over a limited range of factor levels, it has no basis in theory and we cannot reliably extend it to unexplored parts of the response surface.

FACTORIAL DESIGNS

To build an empirical model we measure the response for at least two levels for each factor. For convenience we label these levels as high, H_f and low, L_f where f is the factor; thus H_A is the high level for factor A and L_B is the low level for factor B . If our empirical model contains more than one factor, then each factor's high level is paired with both the high level and the low level for all other factors. In the same way, the low level for each factor is paired with the high level and the low level for all other factors. As shown in Figure 14.12, this requires 2^k experiments where k is the number of factors. This experimental design is known as a **2^k FACTORIAL DESIGN**.

Another system of notation is to use a plus sign (+) to indicate a factor's high level and a minus sign (-) to indicate its low level. We will use H or L when writing an equation and a plus sign or a minus sign in tables.

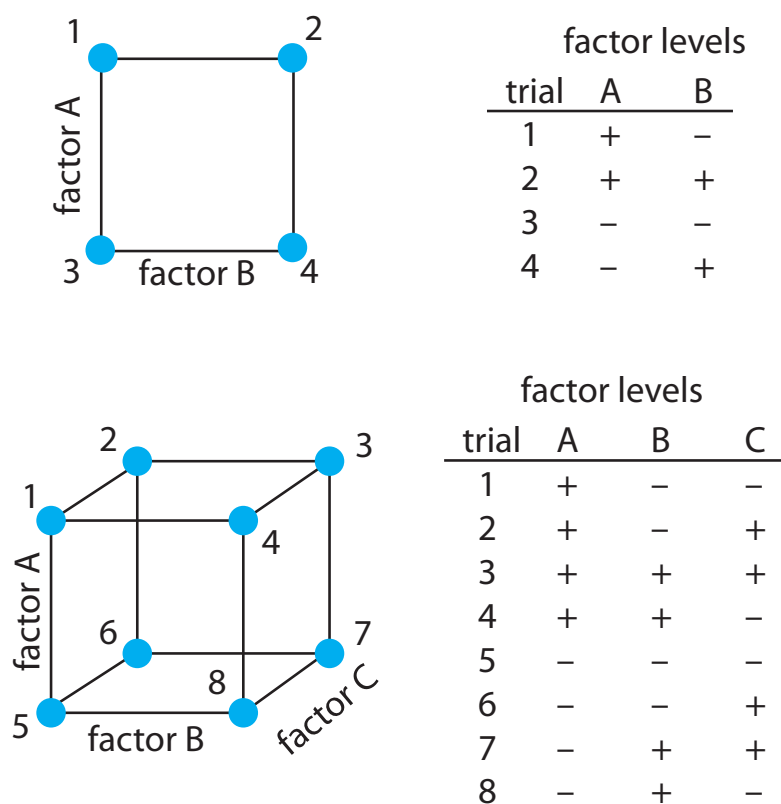


Figure 14.12 2^k factorial designs for (top) $k = 2$, and for (bottom) $k = 3$. A 2^2 factorial design requires four experiments and a 2^3 factorial design requires eight experiments.

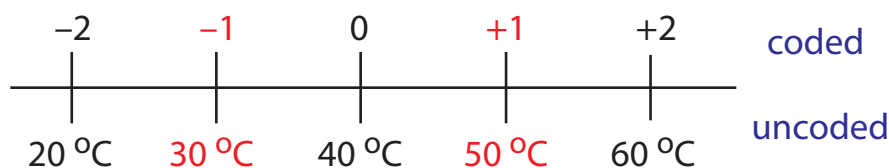


Figure 14.13 The relationship between the coded factor levels and the uncoded factor levels for [Example 14.2](#). The numbers in red are the values defined in the 2^2 factorial design.

CODED FACTOR LEVELS

The calculations for a 2^k factorial design are straightforward and easy to complete with a calculator or a spreadsheet. To simplify the calculations, we code the factor levels using +1 for a high level and -1 for a low level. Coding has two additional advantages: scaling the factors to the same magnitude makes it easier to evaluate each factor's relative importance, and it places the model's intercept, β_0 , at the center of the experimental design. As shown in [Example 14.2](#), it is easy to convert between coded and uncoded factor levels.

Example 14.2

To explore the effect of temperature on a reaction, we assign 30 °C to a coded factor level of -1, and assign a coded level +1 to a temperature of 50 °C. What temperature corresponds to a coded level of -0.5 and what is the coded level for a temperature of 60 °C?

SOLUTION

The difference between -1 and +1 is 2, and the difference between 30 °C and 50 °C is 20 °C; thus, each unit in coded form is equivalent to 10 °C in uncoded form. With this information, it is easy to create a simple scale between the coded and the uncoded values, as shown in [Figure 14.13](#). A temperature of 35 °C corresponds to a coded level of -0.5 and a coded level of +2 corresponds to a temperature of 60 °C.

DETERMINING THE EMPIRICAL MODEL

Let's begin by considering a simple example that involves two factors, A and B , and the following empirical model.

$$R = \beta_0 + \beta_a A + \beta_b B + \beta_{ab} AB \quad 14.5$$

A 2^k factorial design with two factors requires four runs. [Table 14.3](#) provides the uncoded levels (A and B), the coded levels (A^* and B^*), and the responses (R) for these experiments. The terms β_0 , β_a , β_b , and β_{ab} in equation 14.5 account for, respectively, the mean effect (which is the average response), the first-order effects due to factor A and to factor B , and the interaction between the two factors.

Equation 14.5 has four unknowns—the four beta terms—and [Table 14.3](#) describes the four experiments. We have just enough information to

Table 14.3 Example of Uncoded and Coded Factor Levels and Responses for a 2² Factorial Design

run	A	B	A*	B*	R
1	15	30	+1	+1	22.5
2	15	10	+1	-1	11.5
3	5	30	-1	+1	17.5
4	5	10	-1	-1	8.5

calculate values for β_0 , β_a , β_b , and β_{ab} . When working with the coded factor levels, the values of these parameters are easy to calculate using the following equations, where n is the number of runs.

$$\beta_0 \approx b_0 = \frac{1}{n} \sum_{i=1}^n R_i \quad 14.6$$

$$\beta_a \approx b_a = \frac{1}{n} \sum_{i=1}^n A_i^* R_i \quad 14.7$$

$$\beta_b \approx b_b = \frac{1}{n} \sum_{i=1}^n B_i^* R_i \quad 14.8$$

$$\beta_{ab} \approx b_{ab} = \frac{1}{n} \sum_{i=1}^n A_i^* B_i^* R_i \quad 14.9$$

Solving for the estimated parameters using the data in Table 14.3

$$b_0 = \frac{22.5 + 11.5 + 17.5 + 8.5}{4} = 15.0$$

$$b_a = \frac{22.5 + 11.5 - 17.5 - 8.5}{4} = 2.0$$

$$b_b = \frac{22.5 - 11.5 + 17.5 - 8.5}{4} = 5.0$$

$$b_{ab} = \frac{22.5 - 11.5 - 17.5 + 8.5}{4} = 0.5$$

leaves us with the coded empirical model for the response surface.

$$R = 15.0 + 2.0A^* + 5.0B^* + 0.05A^*B^* \quad 14.10$$

We can extend this approach to any number of factors. For a system with three factors— A , B , and C —we can use a 2³ factorial design to determine the parameters in the following empirical model

$$R = \beta_0 + \beta_a A + \beta_b B + \beta_c C + \beta_{ab} AB + \beta_{ac} AC + \beta_{bc} BC + \beta_{abc} ABC \quad 14.11$$

where A , B , and C are the factor levels. The terms β_0 , β_a , β_b , and β_{ab} are estimated using equation 14.6, equation 14.7, equation 14.8, and equation 14.9, respectively. To find estimates for the remaining parameters we use the following equations.

In Section 5D.1 of Chapter 5 we introduced the convention of using β to indicate the true value of a regression's model's parameter's and b to indicate its calculated value. We estimate β from b .

Recall that we introduced coded factor levels with the promise that they simplify calculations.

Although we can convert this coded model into its uncoded form, there is no need to do so. If we need to know the response for a new set of factor levels, we just convert them into coded form and calculate the response. For example, if A is 10 and B is 15, then A^* is 0 and B^* is -0.5. Substituting these values into equation 14.10 gives a response of 12.5.

$$\beta_c \approx b_c = \frac{1}{n} \sum_{i=1}^n C_i^* R_i \quad 14.12$$

$$\beta_{ac} \approx b_{ac} = \frac{1}{n} \sum_{i=1}^n A_i^* C_i^* R_i \quad 14.13$$

$$\beta_{bc} \approx b_{bc} = \frac{1}{n} \sum_{i=1}^n B_i^* C_i^* R_i \quad 14.14$$

$$\beta_{abc} \approx b_{abc} = \frac{1}{n} \sum_{i=1}^n A_i^* B_i^* C_i^* R_i \quad 14.15$$

Example 14.3

Table 14.4 lists the uncoded factor levels, the coded factor levels, and the responses for a 2^3 factorial design. Determine the coded empirical model for the response surface based on [equation 14.11](#). What is the expected response when A is 10, B is 15, and C is 50?

SOLUTION

Equation 14.5 has eight unknowns—the eight beta terms—and Table 14.4 describes eight experiments. We have just enough information to calculate values for β_0 , β_a , β_b , β_c , β_{ab} , β_{ac} , β_{bc} , and β_{abc} ; these values are

$$b_0 = \frac{1}{8} \times (137.25 + 54.75 + 73.75 + 30.25 + 61.75 + 30.25 + 41.25 + 18.75) = 56.0$$

$$b_a = \frac{1}{8} \times (137.25 + 54.75 + 73.75 + 30.25 - 61.75 - 30.25 - 41.25 - 18.75) = 18.0$$

$$b_b = \frac{1}{8} \times (137.25 + 54.75 - 73.75 - 30.25 + 61.75 + 30.25 - 41.25 - 18.75) = 15.0$$

Table 14.4 Example of Uncoded and Coded Factor Levels and Responses for the 2^3 Factorial Design in Example 14.3

run	A	B	C	A^*	B^*	C^*	R
1	15	30	45	+1	+1	+1	137.25
2	15	30	15	+1	+1	-1	54.75
3	15	10	45	+1	-1	+1	73.75
4	15	10	15	+1	-1	-1	30.25
5	5	30	45	-1	+1	+1	61.75
6	5	30	15	-1	+1	-1	30.25
7	5	10	45	-1	-1	+1	41.25
8	5	10	15	-1	-1	-1	18.75

	-2	-1	0	+1	+2	coded
	----- ----- ----- ----- -----					
A	0	5	10	15	20	uncoded
B	0	10	20	30	40	
C	0	15	30	45	60	

Figure 14.14 The relationship between the coded factor levels and the uncoded factor levels for [Example 14.3](#). The numbers in red are the values defined in the 2^3 factorial design.

$$b_c = \frac{1}{8} \times (137.25 - 54.75 + 73.75 - 30.25 + 61.75 - 30.25 + 41.25 - 18.75) = 22.5$$

$$b_{ab} = \frac{1}{8} \times (137.25 + 54.75 - 73.75 - 30.25 - 61.75 - 30.25 + 41.25 + 18.75) = 7.0$$

$$b_{ac} = \frac{1}{8} \times (137.25 - 54.75 + 73.75 - 30.25 - 61.75 + 30.25 - 41.25 + 18.75) = 9.0$$

$$b_{bc} = \frac{1}{8} \times (137.25 - 54.75 - 73.75 + 30.25 + 61.75 - 30.25 - 41.25 + 18.75) = 6.0$$

$$b_{abc} = \frac{1}{8} \times (137.25 - 54.75 - 73.75 + 30.25 - 61.75 + 30.25 + 41.25 - 18.75) = 3.75$$

The coded empirical model, therefore, is

$$R = 56.0 + 18.0A^* + 15.0B^* + 22.5C^* + 7.0A^*B^* + 9.0A^*C^* + 6.0B^*C^* + 3.75A^*B^*C^*$$

To find the response when A is 10, B is 15, and C is 50, we first convert these values into their coded form. [Figure 14.14](#) helps us make the appropriate conversions; thus, A^* is 0, B^* is -0.5 , and C^* is $+1.33$. Substituting back into the empirical model gives a response of

$$R = 56.0 + 18.0(0) + 15.0(-0.5) + 22.5(1.33) + 7.0(0)(-0.5) + 9.0(0)(1.33) + 6.0(-0.5)(1.33) + 3.75(0)(-0.5)(1.33) = 74.435 \approx 74.4$$

A 2^k factorial design can model only a factor's first-order effect, including first-order interactions, on the response. A 2^2 factorial design, for example, includes each factor's first-order effect (β_a and β_b) and a first-order interaction between the factors (β_{ab}). A 2^k factorial design cannot model higher-order effects because there is insufficient information. Here is simple example that illustrates the problem. Suppose we need to model a system in

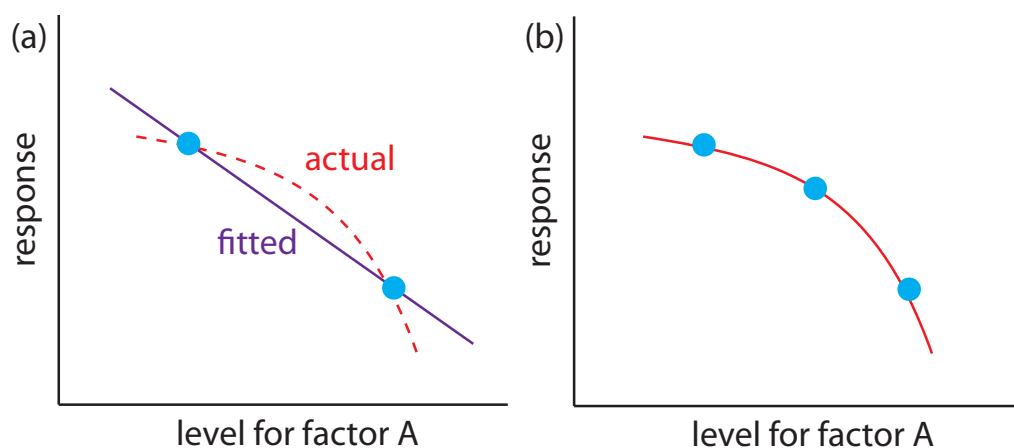


Figure 14.15 A curved one-factor response surface, in red, showing (a) the limitation of using a 2^1 factorial design, which can fit only a straight-line to the data, and (b) the application of a 3^1 factorial design that takes into account second-order effects.

which the response is a function of a single factor, A . Figure 14.15a shows the result of an experiment using a 2^1 factorial design. The only empirical model we can fit to the data is a straight line.

$$R = \beta_0 + \beta_a A$$

If the actual response is a curve instead of a straight-line, then the empirical model is in error. To see evidence of curvature we must measure the response for at least three levels for each factor. We can fit the 3^1 factorial design in Figure 14.15b to an empirical model that includes second-order factor effects.

$$R = \beta_0 + \beta_a A + \beta_{aa} A^2$$

In general, an n -level factorial design can model single-factor and interaction terms up to the $(n-1)$ th order.

We can judge the effectiveness of a first-order empirical model by measuring the response at the center of the factorial design. If there are no higher-order effects, then the average response of the trials in a 2^k factorial design should equal the measured response at the center of the factorial design. To account for influence of random errors we make several determinations of the response at the center of the factorial design and establish a suitable confidence interval. If the difference between the two responses is significant, then a first-order empirical model probably is inappropriate.

One of the advantages of working with a coded empirical model is that b_0 is the average response of the $2 \times k$ trials in a 2^k factorial design.

Example 14.4

One method for the quantitative analysis of vanadium is to acidify the solution by adding H_2SO_4 and oxidizing the vanadium with H_2O_2 to form a red-brown soluble compound with the general formula $(\text{VO})_2(\text{SO}_4)_3$. Palasota and Deming studied the effect of the relative amounts of H_2SO_4

and H_2O_2 on the solution's absorbance, reporting the following results for a 2^2 factorial design.⁵

H_2SO_4	H_2O_2	absorbance
+1	+1	0.330
+1	-1	0.359
-1	+1	0.293
-1	-1	0.420

Four replicate measurements at the center of the factorial design give absorbances of 0.334, 0.336, 0.346, and 0.323. Determine if a first-order empirical model is appropriate for this system. Use a 90% confidence interval when accounting for the effect of random error.

SOLUTION

We begin by determining the confidence interval for the response at the center of the factorial design. The mean response is 0.335 with a standard deviation of 0.0094, which gives a 90% confidence interval of

$$\mu = \bar{X} \pm \frac{ts}{\sqrt{n}} = 0.335 \pm \frac{(2.35)(0.0094)}{\sqrt{4}} = 0.335 \pm 0.011$$

The average response, \bar{R} , from the factorial design is

$$\bar{R} = \frac{0.330 + 0.359 + 0.293 + 0.420}{4} = 0.350$$

Because \bar{R} exceeds the confidence interval's upper limit of 0.346, we can reasonably assume that a 2^2 factorial design and a first-order empirical model are inappropriate for this system at the 95% confidence level.

If we cannot fit a first-order empirical model to our data, we may be able to model it using a full second-order polynomial equation, such as that shown here for a two factors.

$$R = \beta_0 + \beta_a A + \beta_b B + \beta_{aa} A^2 + \beta_{bb} B^2 + \beta_{ab} AB$$

Because we must measure each factor for at least three levels if we are to detect curvature (see Figure 14.15b), a convenient experimental design is a 3^k factorial design. A 3^2 factorial design for two factors, for example, is shown in Figure 14.16. The computations for 3^k factorial designs are not as easy to generalize as those for a 2^k factorial design and are not considered in this text. See this chapter's additional resources for details about the calculations.

CENTRAL COMPOSITE DESIGNS

One limitation to a 3^k factorial design is the number of trials we need to run. As shown in Figure 14.16, a 3^2 factorial design requires 9 trials. This number increases to 27 for three factors and to 81 for 4 factors. A more efficient

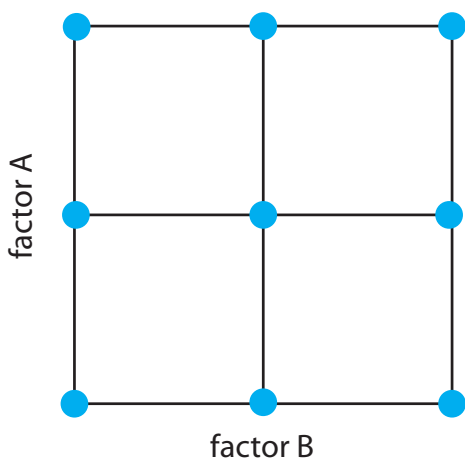


Figure 14.16 A 3^k factorial design for $k=2$.

Problem 14.11 in the end-of-chapter problems provides a complete empirical model for this system.

5 Palasota, J. A.; Deming, S. N. *J. Chem. Educ.* **1992**, *62*, 560–563.

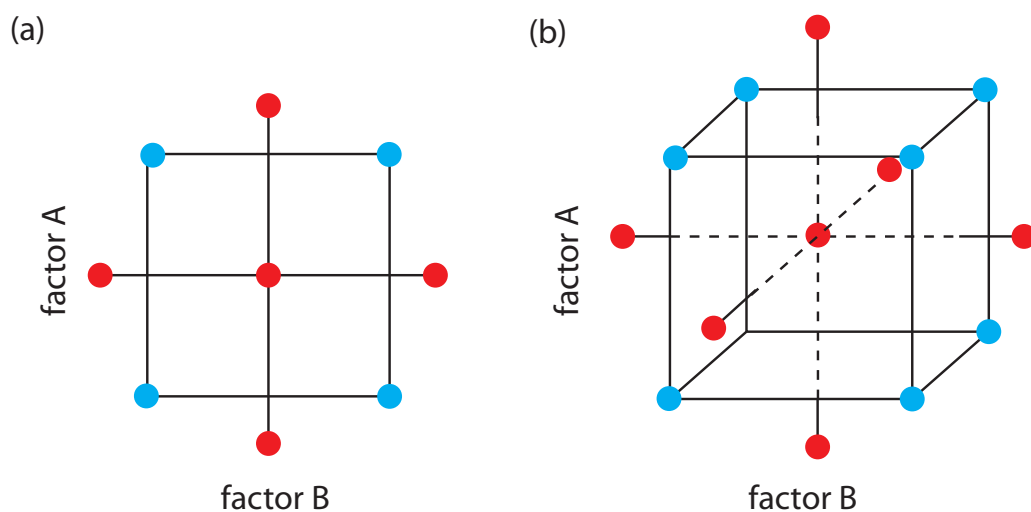


Figure 14.17 Two examples of a central composite design for (a) $k=2$ and (b) $k=3$. The points in **blue** are a 2^k factorial design, and the points in **red** are a star design.

experimental design for a system that contains more than two factors is a **CENTRAL COMPOSITE DESIGN**, two examples of which are shown in Figure 14.17. The central composite design consists of a 2^k factorial design, which provides data to estimate each factor's first-order effect and interactions between the factors, and a star design that has $2k + 1$ points, which provides data to estimate second-order effects. Although a central composite design for two factors requires the same number of trials, nine, as a 3^2 factorial design, it requires only 15 trials and 25 trials when using three factors or four factors. See this chapter's additional resources for details about the central composite designs.

14B Verifying the Method

After developing and optimizing a method, the next step is to determine how well it works in the hands of a single analyst. Three steps make up this process: determining single-operator characteristics, completing a blind analysis of standards, and determining the method's ruggedness. If another standard method is available, then we can analyze the same sample using both the standard method and the new method, and compare the results. If the result for any single test is unacceptable, then the method is not a suitable standard method.

14B.1 Single Operator Characteristics

The first step in verifying a method is to determine the precision, accuracy, and detection limit when a single analyst uses the method to analyze a standard sample. The detection limit is determined by analyzing an appropriate reagent blank. Precision is determined by analyzing replicate portions of the sample, preferably more than ten. Accuracy is evaluated using a t -test

See Chapter 4G for a discussion of detection limits. Pay particular attention to the difference between a detection limit, a limit of identification, and a limit of quantitation.

See Section 4F.1 for a review of the *t*-test.

See Chapter 4B for a review of constant determinate errors. Figure 4.5 illustrates how we can detect a constant determinate error by analyzing samples containing different amounts of analyte.

An even more stringent requirement is to require that the experimental mean be within two standard deviations of the analyte's known concentration.

For example, if temperature is a concern, we might specify that it be held at $25 \pm 2^\circ\text{C}$.

to compare the experimental results to the known amount of analyte in the standard. Precision and accuracy are evaluated for several different concentrations of analyte, including at least one concentration near the detection limit, and for each different sample matrix. Including different concentrations of analyte helps to identify constant sources of determinate error and to establish the range of concentrations for which the method is applicable.

14B.2 Blind Analysis of Standard Samples

Single-operator characteristics are determined by analyzing a standard sample that has a concentration of analyte known to the analyst. The second step in verifying a method is a **BLIND ANALYSIS** of standard samples. Although the concentration of analyte in the standard is known to a supervisor, the information is withheld from the analyst. After analyzing the standard sample several times, the analyte's average concentration is reported to the test's supervisor. To be accepted, the experimental mean must be within three standard deviations—as determined from the single-operator characteristics—of the analyte's known concentration.

14B.3 Ruggedness Testing

An optimized method may produce excellent results in the laboratory that develops a method, but poor results in other laboratories. This is not particularly surprising because a method typically is optimized by a single analyst using the same reagents, equipment, and instrumentation for each trial. Any variability introduced by different analysts, reagents, equipment, and instrumentation is not included in the single-operator characteristics. Other less obvious factors may affect an analysis, including environmental factors, such as the temperature or relative humidity in the laboratory; if the procedure does not require control of these conditions, then they may contribute to variability. Finally, the analyst who optimizes the method usually takes particular care to perform the analysis in exactly the same way during every trial, which may minimize the run-to-run variability.

An important step in developing a standard method is to determine which factors have a pronounced effect on the quality of the results. Once we identify these factors, we can write specific instructions that specify how these factors must be controlled. A procedure that, when carefully followed, produces results of high quality in different laboratories is considered **RUGGEDNESS TESTING**.⁶

Ruggedness testing usually is performed by the laboratory that develops the standard method. After identifying potential factors, their effects on the response are evaluated by performing the analysis at two levels for each factor. Normally one level is that specified in the procedure, and the other is a level likely encountered when the procedure is used by other laboratories.

⁶ Youden, W. J. *Anal. Chem.* **1960**, *32*(13), 23A–37A.

Table 14.5 Experimental Design for a Ruggedness Test Involving Seven Factors

run	factors							response
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	
1	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	R_1
2	<i>A</i>	<i>B</i>	<i>c</i>	<i>D</i>	<i>e</i>	<i>f</i>	<i>g</i>	R_2
3	<i>A</i>	<i>b</i>	<i>C</i>	<i>d</i>	<i>E</i>	<i>f</i>	<i>g</i>	R_3
4	<i>A</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>F</i>	<i>G</i>	R_4
5	<i>a</i>	<i>B</i>	<i>C</i>	<i>d</i>	<i>e</i>	<i>F</i>	<i>g</i>	R_5
6	<i>a</i>	<i>B</i>	<i>c</i>	<i>d</i>	<i>E</i>	<i>f</i>	<i>G</i>	R_6
7	<i>a</i>	<i>b</i>	<i>C</i>	<i>D</i>	<i>e</i>	<i>f</i>	<i>G</i>	R_7
8	<i>a</i>	<i>b</i>	<i>c</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>g</i>	R_8

This approach to ruggedness testing can be time consuming. If there are seven potential factors, for example, a 2^7 factorial design can evaluate each factor's first-order effect. Unfortunately, this requires a total of 128 trials—too many trials to be a practical solution. A simpler experimental design is shown in Table 14.5, in which the two factor levels are identified by upper case and lower case letters. This design, which is similar to a 2^3 factorial design, is called a fractional factorial design. Because it includes only eight runs, the design provides information only the average response and the seven first-order factor effects. It does not provide sufficient information to evaluate higher-order effects or interactions between factors, both of which are probably less important than the first-order effects.

The experimental design in Table 14.5 is balanced in that each of a factor's two levels is paired an equal number of times with the upper case and lower case levels for every other factor. To determine the effect, E , of changing a factor's level, we subtract the average response when the factor is at its upper case level from the average value when it is at its lower case level.

$$E = \frac{(\sum R_i)_{\text{upper case}}}{4} - \frac{(\sum R_i)_{\text{lower case}}}{4} \quad 14.16$$

Because the design is balanced, the levels for the remaining factors appear an equal number of times in both summation terms, canceling their effect on E . For example, to determine the effect of factor A , E_A , we subtract the average response for runs 5–8 from the average response for runs 1–4. Factor B does not affect E_A because its upper case levels in runs 1 and 2 are canceled by the upper case levels in runs 5 and 6, and its lower case levels in runs 3 and 4 are canceled by the lower case levels in runs 7 and 8. After we calculate each of the factor effects we rank them from largest to smallest without regard to sign, identifying those factors whose effects are substantially larger than the other factors.

Why does this model estimate the seven first-order factor effects, E , and not seven of the 20 possible first-order interactions? With eight experiments, we can only choose to calculate seven parameters (plus the average response). The calculation of E_D , for example, also gives the value for E_{AB} . You can convince yourself of this by replacing each upper case letter with a +1 and each lower case letter with a -1 and noting that $A \times B = D$. We choose to report the first-order factor effects because they likely are more important than interactions between factors.

To see that this design is balanced, look closely at the last four runs. Factor A is present at its level a for all four of these runs. For each of the remaining factors, two levels are upper case and two levels are lower case. Runs 5–8 provide information about the effect of a on the response, but do not provide information about the effect of any other factor. Runs 1, 2, 5, and 6 provide information about the effect of B , but not of the remaining factors. Try a few other examples to convince yourself that this relationship is general.

We also can use this experimental design to estimate the method's expected standard deviation due to the effects of small changes in uncontrolled or poorly controlled factors.⁷

$$s = \sqrt{\frac{2}{7} \sum_{i=1}^n E_i^2} \quad 14.17$$

If this standard deviation is too large, then the procedure is modified to bring under control the factors that have the greatest effect on the response.

Example 14.5

The concentration of trace metals in sediment samples collected from rivers and lakes are determined by extracting with acid and analyzing the extract by atomic absorption spectrophotometry. One procedure calls for an overnight extraction using dilute HCl or HNO₃. The samples are placed in plastic bottles with 25 mL of acid and then placed on a shaker operated at a moderate speed and at ambient temperature. To determine the method's ruggedness, the effect of the following factors was studied using the experimental design in [Table 14.5](#).

Factor A: extraction time	$A = 24 \text{ h}$	$a = 12 \text{ h}$
Factor B: shaking speed	$B = \text{medium}$	$b = \text{high}$
Factor C: acid type	$C = \text{HCl}$	$c = \text{HNO}_3$
Factor D: acid concentration	$D = 0.1 \text{ M}$	$d = 0.05 \text{ M}$
Factor E: volume of acid	$E = 25 \text{ mL}$	$e = 35 \text{ mL}$
Factor F: type of container	$F = \text{plastic}$	$f = \text{glass}$
Factor G: temperature	$G = \text{ambient}$	$g = 25^\circ\text{C}$

Eight replicates of a standard sample that contains a known amount of analyte are carried through the procedure. The percentage of analyte recovered in the eight samples are as follows: $R_1 = 98.9$, $R_2 = 99.0$, $R_3 = 97.5$, $R_4 = 97.7$, $R_5 = 97.4$, $R_6 = 97.3$, $R_7 = 98.6$, and $R_8 = 98.6$. Identify the factors that have a significant effect on the response and estimate the method's expected standard deviation.

SOLUTION

To calculate the effect of changing each factor's level we use [equation 14.16](#) and substitute in appropriate values. For example, E_A is

$$E_A = \frac{98.9 + 99.0 + 97.5 + 97.7}{4} - \frac{97.4 + 97.3 + 98.6 + 98.6}{4} = 0.30$$

⁷ Youden, W. J. "Statistical Techniques for Collaborative Tests," in *Statistical Manual of the Association of Official Analytical Chemists*, Association of Official Analytical Chemists: Washington, D. C., 1975, p. 35.

Completing the remaining calculations and ordering the factors by the absolute values of their effects

Factor <i>D</i>	1.30
Factor <i>A</i>	0.35
Factor <i>E</i>	-0.10
Factor <i>B</i>	0.05
Factor <i>C</i>	-0.05
Factor <i>F</i>	0.05
Factor <i>G</i>	0.00

shows us that the concentration of acid (Factor *D*) has a substantial effect on the response, with a concentration of 0.05 M providing a much lower percent recovery. The extraction time (Factor *A*) also appears significant, but its effect is not as important as the acid's concentration. All other factors appear insignificant. The method's estimated standard deviation is

$$s = \sqrt{\frac{2}{7} \left\{ (1.30)^2 + (0.35)^2 + (-0.10)^2 + (0.05)^2 + (-0.05)^2 + (0.05)^2 + (0.00)^2 \right\}} = 0.72$$

which, for an average recovery of 98.1% gives a relative standard deviation of approximately 0.7%. If we control the acid's concentration so that its effect approaches that for factors *B*, *C*, and *F*, then the relative standard deviation becomes 0.18, or approximately 0.2%.

14B.4 Equivalency Testing

If an approved standard method is available, then a new method should be evaluated by comparing results to those obtained when using the standard method. Normally this comparison is made at a minimum of three concentrations of analyte to evaluate the new method over a wide dynamic range. Alternatively, we can plot the results obtained using the new method against results obtained using the approved standard method. A slope of 1.00 and a *y*-intercept of 0.0 provides evidence that the two methods are equivalent.

14C Validating the Method as a Standard Method

For an analytical method to be useful, an analyst must be able to achieve results of acceptable accuracy and precision. Verifying a method, as described in the previous section, establishes this goal for a single analyst. Another requirement for a useful analytical method is that an analyst should obtain the same result from day-to-day, and different labs should obtain the same result when analyzing the same sample. The process by which we approve a method for general use is known as **VALIDATION** and it involves a collaborative test of the method by analysts in several laboratories. Collaborative testing is used routinely by regulatory agencies and professional organizations,

Representative Method 10.1 for the determination of iron in water and wastewater, and Representative Method 10.5 for the determination of sulfate in water, are two examples of standard methods validated through collaborative testing.

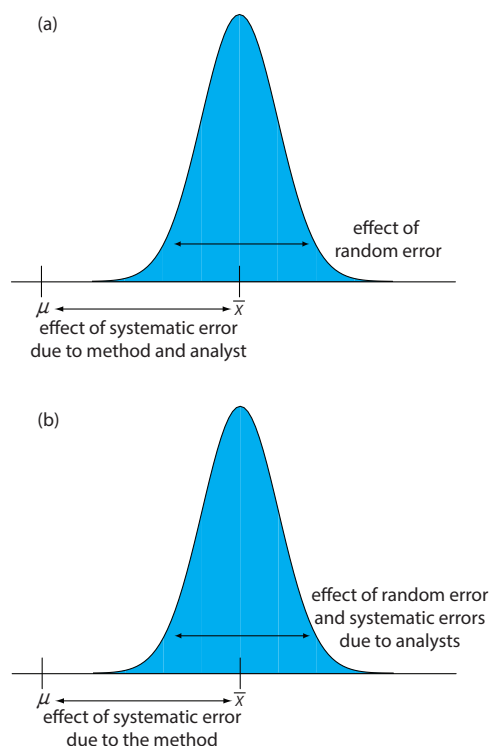


Figure 14.18 Partitioning of random errors, systematic errors due to the analyst, and systematic errors due to the method for (a) replicate analyses performed by a single analyst and (b) single determinations performed by several analysts.

such as the U. S. Environmental Protection Agency, the American Society for Testing and Materials, the Association of Official Analytical Chemists, and the American Public Health Association. Many of the representative methods in earlier chapters are identified by these agencies as validated methods.

When an analyst performs a single analysis on a single sample the difference between the experimentally determined value and the expected value is influenced by three sources of error: random errors, systematic errors inherent to the method, and systematic errors unique to the analyst. If the analyst performs enough replicate analyses, then we can plot a distribution of results, as shown in Figure 14.18a. The width of this distribution is described by a standard deviation that provides an estimate of the random errors affecting the analysis. The position of the distribution's mean, \bar{X} , relative to the sample's true value, μ , is determined both by systematic errors inherent to the method and those systematic errors unique to the analyst. For a single analyst there is no way to separate the total systematic error into its component parts.

The goal of a **COLLABORATIVE TEST** is to determine the magnitude of all three sources of error. If several analysts each analyze the same sample one time, the variation in their collective results (see Figure 14.18b) includes contributions from random errors and systematic errors (biases) unique to the analysts. Without additional information, we cannot separate the standard deviation for this pooled data into the precision of the analysis and the systematic errors introduced by the analysts. We can use the position of the distribution, to detect the presence of a systematic error in the method.

14C.1 Two-Sample Collaborative Testing

The design of a collaborative test must provide the additional information needed to separate random errors from the systematic errors introduced by the analysts. One simple approach—accepted by the Association of Official Analytical Chemists—is to have each analyst analyze two samples that are similar in both their matrix and in their concentration of analyte. To analyze the results we represent each analyst as a single point on a two-sample scatterplot, using the result for one sample as the x -coordinate and the result for the other sample as the y -coordinate.⁸

As shown in Figure 14.19, a two-sample chart places each analyst into one of four quadrants, which we identify as (+, +), (−, +), (−, −) and (+, −). A plus sign indicates the analyst's result for a sample is greater than the mean for all analysts and a minus sign indicates the analyst's result is less than the mean for all analysts. The quadrant (+, −), for example, contains those analysts that exceeded the mean for sample X and that undershot the mean for sample Y . If the variation in results is dominated by random errors, then

⁸ Youden, W. J. "Statistical Techniques for Collaborative Tests," in *Statistical Manual of the Association of Official Analytical Chemists*, Association of Official Analytical Chemists: Washington, D. C., 1975, pp 10–11.

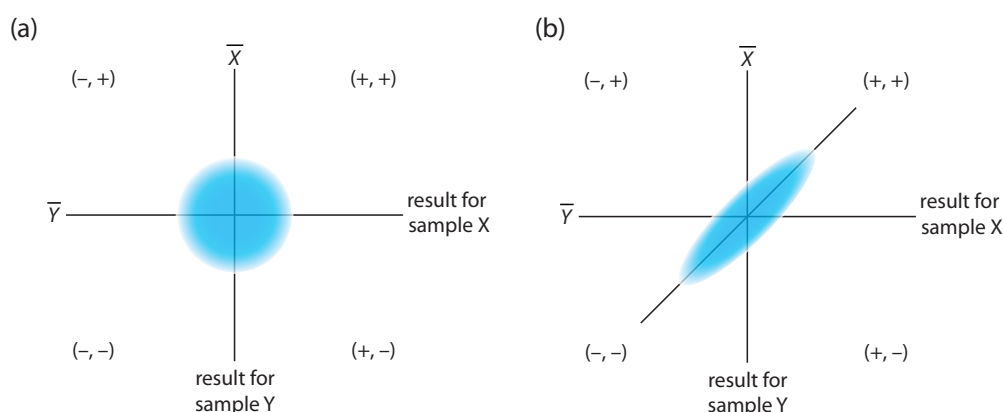


Figure 14.19 Typical two-sample plots when (a) random errors are significantly larger than systematic errors due to the analysts, and (b) when systematic errors due to the analysts are significantly larger than the random errors.

we expect the points to be distributed randomly in all four quadrants, with an equal number of points in each quadrant. Furthermore, as shown in Figure 14.19a, the points will cluster in a circular pattern whose center is the mean values for the two samples. When systematic errors are significantly larger than random errors, then the points fall primarily in the (+, +) and the (-, -) quadrants, forming an elliptical pattern around a line that bisects these quadrants at a 45° angle, as seen in Figure 14.19b.

A visual inspection of a two-sample chart is an effective method for qualitatively evaluating the capabilities of a proposed standard method, as shown in figure 14.20. The length of a perpendicular line from any point to the 45° line is proportional to the effect of random error on that analyst's results. The distance from the intersection of the axes—which corresponds to the mean values for samples X and Y—to the perpendicular projection of a point on the 45° line is proportional to the analyst's systematic error. An ideal standard method has small random errors and small systematic errors due to the analysts, and has a compact clustering of points that is more circular than elliptical.

We also can use the data in a two-sample chart to separate the total variation in the data, σ_{tot} , into contributions from random error, σ_{rand} , and from systematic errors due to the analysts, σ_{syst} .⁹ Because an analyst's systematic errors are present in his or her analysis of both samples, the difference, D , between the results estimates the contribution of random error.

$$D_i = X_i - Y_i$$

To estimate the total contribution from random error we use the standard deviation of these differences, s_D , for all analysts

$$s_D = \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{2(n-1)}} = s_{\text{rand}} \approx \sigma_{\text{rand}} \quad 14.18$$

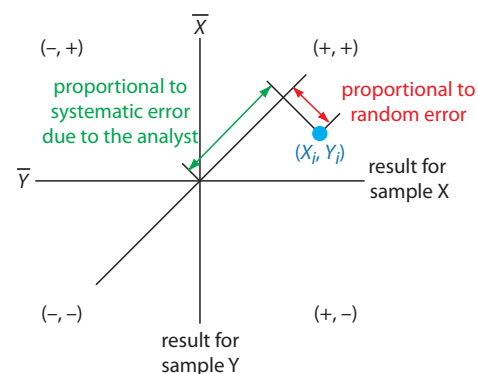


Figure 14.20 Relationship between the result for a single analyst (in blue) and the contribution of random error (red arrow) and the contribution from the analyst's systematic error (green arrow).

⁹ Youden, W. J. "Statistical Techniques for Collaborative Tests," in *Statistical Manual of the Association of Official Analytical Chemists*, Association of Official Analytical Chemists: Washington, D. C., 1975, pp 22–24.

where n is the number of analysts. The factor of 2 in the denominator of equation 14.18 is the result of using two values to determine D_i . The total, T_i , of each analyst's results

$$T_i = X_i + Y_i$$

contains contributions from both random error and twice the analyst's systematic error.

$$\sigma_{\text{tot}}^2 = \sigma_{\text{rand}}^2 + 2\sigma_{\text{syst}}^2 \quad 14.19$$

The standard deviation of the totals, s_T , provides an estimate for σ_{tot} .

$$s_T = \sqrt{\frac{\sum_{i=1}^n (T_i - \bar{T})^2}{2(n-1)}} = s_{\text{tot}} \approx \sigma_{\text{tot}} \quad 14.20$$

We double the analyst's systematic error in equation 14.19 because it is the same in each analysis.

Again, the factor of 2 in the denominator is the result of using two values to determine T_i .

If the systematic errors are significantly larger than the random errors, then s_T is larger than s_D , a hypothesis we can evaluate using a one-tailed F -test

$$F = \frac{s_T^2}{s_D^2}$$

where the degrees of freedom for both the numerator and the denominator are $n-1$. As shown in the following example, if s_T is significantly larger than s_D we can use equation 14.19 to separate σ_{tot}^2 into components that represent the random error and the systematic error.

For a review of the F -test, see Section 4F.2 and Section 4F.3. Example 4.18 illustrates a typical application.

Example 14.6

As part of a collaborative study of a new method for determining the amount of total cholesterol in blood, you send two samples to 10 analysts with instructions that they analyze each sample one time. The following results, in mg total cholesterol per 100 mL of serum, are returned to you.

analyst	sample 1	sample 2
1	245.0	229.4
2	247.4	249.7
3	246.0	240.4
4	244.9	235.5
5	255.7	261.7
6	248.0	239.4
7	249.2	255.5
8	225.1	224.3
9	255.0	246.3
10	243.1	253.1

Use this data estimate σ_{rand} and σ_{syst} for the method.

SOLUTION

Figure 14.21 provides a two-sample plot of the results. The clustering of points suggests that the systematic errors of the analysts are significant. The vertical line at 245.9 mg/100 mL is the average value for sample 1 and the average value for sample 2 is indicated by the horizontal line at 243.5 mg/100 mL. To estimate σ_{rand} and σ_{syst} we first calculate values for D_i and T_i .

analyst	D_i	T_i
1	15.6	474.4
2	-2.3	497.1
3	5.6	486.4
4	9.4	480.4
5	-6.0	517.4
6	8.6	487.4
7	-6.3	504.7
8	0.8	449.4
9	8.7	501.3
10	-10.0	496.2

Next, we calculate the standard deviations for the differences, s_D , and the totals, s_T , using equations 14.18 and 14.20, obtaining $s_D = 5.95$ and $s_T = 13.3$. To determine if the systematic errors between the analysts are significant, we use an F -test to compare s_T and s_D .

$$F = \frac{s_T^2}{s_D^2} = \frac{(13.3)^2}{(5.95)^2} = 5.00$$

Because the F -ratio is larger than $F(0.05, 9, 9)$, which is 3.179, we conclude that the systematic errors between the analysts are significant at the 95% confidence level. The estimated precision for a single analyst is

$$\sigma_{\text{rand}} \approx s_{\text{rand}} = s_D = 5.95$$

The estimated standard deviation due to systematic errors between analysts is calculated from [equation 14.19](#).

$$\sigma_{\text{syst}} = \sqrt{\frac{\sigma_{\text{tot}}^2 - \sigma_{\text{rand}}^2}{2}} \approx \sqrt{\frac{s_T^2 - s_D^2}{2}} = \sqrt{\frac{(13.3)^2 - (5.95)^2}{2}} = 8.41$$

If the true values for the two samples are known, we also can test for the presence of a systematic error in the method. If there are no systematic method errors, then the sum of the true values, μ_{tot} , for samples X and Y

$$\mu_{\text{tot}} = \mu_X + \mu_Y$$

should fall within the confidence interval around \bar{T} . We can use a two-tailed t -test of the following null and alternate hypotheses

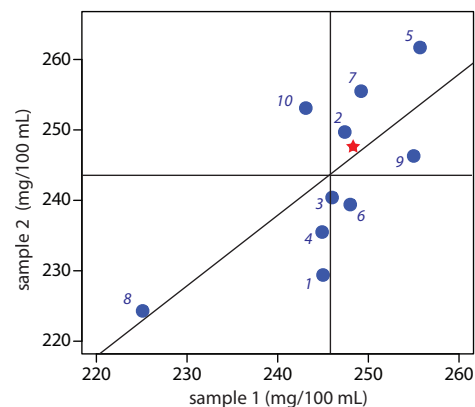


Figure 14.21 Two-sample plot for the data in Example 14.6. The number by each blue point indicates the analyst. The true values for each sample (see [Example 14.7](#)) are indicated by the red star.

Critical values for the F -test are in Appendix 5.

For a review of the t -test of an experimental mean to a known mean, see Section 4F.1. Example 4.16 illustrates a typical application.

$$H_0: \bar{T} = \mu_{\text{tot}} \quad H_A: \bar{T} \neq \mu_{\text{tot}}$$

to determine if there is evidence for a systematic error in the method. The test statistic, t_{exp} , is

$$t_{\text{exp}} = \frac{|\bar{T} - \mu_{\text{tot}}| \sqrt{n}}{s_T \sqrt{2}} \quad 14.21$$

with $n-1$ degrees of freedom. We include the $\sqrt{2}$ in the denominator because s_T (see [equation 14.20](#)) underestimates the standard deviation when comparing \bar{T} to μ_{tot} .

Example 14.7

The two samples analyzed in [Example 14.6](#) are known to contain the following concentrations of cholesterol: $\mu_{\text{samp 1}} = 248.3$ mg/100 mL and $\mu_{\text{samp 2}} = 247.6$ mg/100 mL. Determine if there is any evidence for a systematic error in the method at the 95% confidence level.

SOLUTION

Using the data from [Example 14.6](#) and the true values for the samples, we know that s_T is 13.3, and that

$$\bar{T} = \bar{X}_{\text{samp 1}} + \bar{X}_{\text{samp 2}} = 245.9 + 243.5 = 489.4 \text{ mg/100 mL}$$

$$\mu_{\text{tot}} = \mu_{\text{samp 1}} + \mu_{\text{samp 2}} = 248.3 + 247.6 = 495.9 \text{ mg/100 mL}$$

Substituting these values into equation 14.21 gives

$$t_{\text{exp}} = \frac{|489.4 - 495.9| \sqrt{10}}{13.3 \sqrt{2}} = 1.09$$

Because this value for t_{exp} is smaller than the critical value of 2.26 for $t(0.05, 9)$, there is no evidence for a systematic error in the method at the 95% confidence level.

[Example 14.6](#) and Example 14.7 illustrate how we can use a pair of similar samples in a collaborative test of a new method. Ideally, a collaborative test involves several pairs of samples that span the range of analyte concentrations for which we plan to use the method. In doing so, we evaluate the method for constant sources of error and establish the expected relative standard deviation and bias for different levels of analyte.

14C.2 Collaborative Testing and Analysis of Variance

In a two-sample collaborative test we ask each analyst to perform a single determination on each of two separate samples. After reducing the data to a set of differences, D , and a set of totals, T , each characterized by a mean and a standard deviation, we extract values for the random errors that affect precision and the systematic differences between the analysts. The calculations are relatively simple and straightforward.

Critical values for the t -test are in Appendix 4.

An alternative approach to a collaborative test is to have each analyst perform several replicate determinations on a single, common sample. This approach generates a separate data set for each analyst and requires a different statistical treatment to provide estimates for σ_{rand} and for σ_{sys} .

There are several statistical methods for comparing three or more sets of data. The approach we consider in this section is an **ANALYSIS OF VARIANCE** (ANOVA). In its simplest form, a one-way ANOVA allows us to explore the importance of a single variable—the identity of the analyst is one example—on the total variance. To evaluate the importance of this variable, we compare its variance to the variance explained by indeterminate sources of error.

We first introduced variance in Chapter 4 as one measure of a data set's spread around its central tendency. In the context of an analysis of variance, it is useful for us to understand that variance is simply a ratio of two terms: a sum of squares for the differences between individual values and their mean, and the degrees of freedom. For example, the variance, s^2 , of a data set consisting of n measurements is

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

where X_i is the value of a single measurement and \bar{X} is the mean. The ability to partition the variance into a sum of squares and the degrees of freedom greatly simplifies the calculations in a one-way ANOVA.

Let's use a simple example to develop the rationale behind a one-way ANOVA calculation. The data in Table 14.6 are from four analysts, each asked to determine the purity of a single pharmaceutical preparation of sulfanilamide. Each column in Table 14.6 provides the results for an individual analyst. To help us keep track of this data, we will represent each result as X_{ij} , where i identifies the analyst and j indicates the replicate. For example, $X_{3,5}$ is the fifth replicate for the third analyst, or 94.24%.

The data in Table 14.6 show variability in the results obtained by each analyst and in the difference in the results between the analysts. There are two sources for this variability: indeterminate errors associated with the analytical procedure that are experienced equally by each analyst, and systematic or determinate errors introduced by the individual analysts.

One way to view the data in Table 14.6 is to treat it as a single large sample, characterized by a global mean and a global variance

$$\bar{\bar{X}} = \frac{\sum_{i=1}^b \sum_{j=1}^{n_i} X_{ij}}{N} \quad 14.22$$

$$\bar{\bar{s}}^2 = \frac{\sum_{i=1}^b \sum_{j=1}^{n_i} (X_{ij} - \bar{\bar{X}})^2}{N - 1} \quad 14.23$$

Table 14.6 Determination of the %Purity of a Sulfanilamide Preparation by Four Analysts

replicate	analyst A	analyst B	analyst C	analyst D
1	94.09	99.55	95.14	93.88
2	94.64	98.24	94.62	94.23
3	95.08	101.1	95.28	96.05
4	94.54	100.4	94.59	93.89
5	95.38	100.1	94.24	94.95
6	93.62			95.49
\bar{X}	94.56	99.88	94.77	94.75
s	0.641	1.073	0.428	0.899

where h is the number of samples (in this case the number of analysts), n_i is the number of replicates for the i th sample (in this case the i th analyst), and N is the total number of data points (in this case 22). The global variance—which includes all sources of variability that affect the data—provides an estimate of the combined influence of indeterminate errors and systematic errors.

A second way to work with the data in Table 14.6 is to treat the results for each analyst separately. If we assume that each analyst experiences the same indeterminate errors, then the variance, s^2 , for each analyst provides a separate estimate of σ_{rand}^2 . To pool these individual variances, which we call the **WITHIN-SAMPLE VARIANCE**, s_w^2 , we square the difference between each replicate and its corresponding mean, add them up, and divide by the degrees of freedom.

$$\sigma_{\text{rand}}^2 \approx s_w^2 = \frac{\sum_{i=1}^h \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{N - h} \quad 14.24$$

To estimate the systematic errors, σ_{sys}^2 , that affect the results in Table 14.6 we need to consider the differences between the analysts. The variance of the individual mean values about the global mean, which we call the **BETWEEN-SAMPLE VARIANCE**, s_b^2 , is

$$s_b^2 = \frac{\sum_{i=1}^h n_i (\bar{X}_i - \bar{\bar{X}})^2}{h - 1} \quad 14.25$$

The between-sample variance includes contributions from both indeterminate errors and systematic errors; thus

$$s_b^2 = \sigma_{\text{rand}}^2 + \bar{n} \sigma_{\text{sys}}^2 \quad 14.26$$

where \bar{n} is the average number of replicates per analyst.

$$\bar{n} = \frac{\sum_{i=1}^h n_i}{h}$$

Carefully compare our description of equation 14.24 to the equation itself. It is important that you understand why equation 14.24 provides our best estimate of the indeterminate errors that affect the data in Table 14.6. Note that we lose one degree of freedom for each of the h means included in the calculation.

We lose one degree of freedom for the global mean.

Note the similarity between equation 14.26 and equation 14.19. The analysis of the data in a two-sample plot is the same as a one-way analysis of variance with $h=2$.

In a one-way ANOVA of the data in [Table 14.6](#) we make the null hypothesis that there are no significant differences between the mean values for the analysts. The alternative hypothesis is that at least one of the mean values is significantly different. If the null hypothesis is true, then σ_{sys}^2 must be zero and s_w^2 and s_b^2 should have similar values. If s_b^2 is significantly greater than s_w^2 , then σ_{sys}^2 is greater than zero. In this case we must accept the alternative hypothesis that there is a significant difference between the means for the analysts. The test statistic is the F -ratio

$$F_{\text{exp}} = \frac{s_b^2}{s_w^2}$$

which is compared to the critical value $F(\alpha, b-1, N-b)$. This is a one-tailed significance test because we are interested only in whether s_b^2 is significantly greater than s_w^2 .

Both s_b^2 and s_w^2 are easy to calculate for small data sets. For larger data sets, calculating s_w^2 is tedious. We can simplify the calculations by taking advantage of the relationship between the sum-of-squares terms for the global variance ([equation 14.23](#)), the within-sample variance ([equation 14.24](#)), and the between-sample variance ([equation 14.25](#)). We can split the numerator of [equation 14.23](#), which is the total sum-of-squares, SS_t , into two terms

$$SS_t = SS_w + SS_b$$

where SS_w is the sum-of-squares for the within-sample variance and SS_b is the sum-of-squares for the between-sample variance. Calculating SS_t and SS_b gives SS_w by difference. Finally, dividing SS_w and SS_b by their respective degrees of freedom gives s_w^2 and s_b^2 . [Table 14.7](#) summarizes the equations for a one-way ANOVA calculation. [Example 14.8](#) walks you through the calculations, using the data in [Table 14.6](#). Section 14E provides instructions on using Excel and R to complete a one-way analysis of variance.

[Problem 14.17](#) in the end of chapter problems asks you to verify this relationship between the sum-of-squares.

Table 14.7 Summary of Calculations for a One-Way Analysis of Variance

source	sum-of-squares	degrees of freedom	variance	expected variance	F -ratio
between samples	$SS_b = \sum_{i=1}^h n_i (\bar{X}_i - \bar{X})^2$	$h-1$	$s_b^2 = \frac{SS_b}{h-1}$	$s_b^2 = \sigma_{\text{rand}}^2 + \bar{n}\sigma_{\text{sys}}^2$	$F_{\text{exp}} = \frac{s_b^2}{s_w^2}$
within samples	$SS_w = SS_t - SS_b$	$N-h$	$s_w^2 = \frac{SS_w}{N-h}$	$s_w^2 = \sigma_{\text{rand}}^2$	
total	$SS_t = \sum_{i=1}^h \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$ $= \bar{s}^2 (N-1)$	$N-1$			

Example 14.8

The data in [Table 14.6](#) are from four analysts, each asked to determine the purity of a single pharmaceutical preparation of sulfanilamide. Determine if the difference in their results is significant at $\alpha = 0.05$. If such a difference exists, estimate values for σ_{sys}^2 and σ_{rand}^2 .

SOLUTION

To begin we calculate the global mean ([equation 14.22](#)) and the global variance ([equation 14.23](#)) for the pooled data, and the means for each analyst; these values are summarized here.

$$\begin{aligned}\bar{X} &= 95.87 & \bar{s}^2 &= 5.506 \\ \bar{X}_A &= 94.56 & \bar{X}_B &= 99.88 & \bar{X}_C &= 94.77 & \bar{X}_D &= 94.75\end{aligned}$$

Using these values we calculate the total sum of squares

$$SS_t = \bar{s}^2(N - 1) = (5.506)(22 - 1) = 115.63$$

the between sample sum of squares

$$\begin{aligned}SS_b &= \sum_{i=1}^h n_i(\bar{X}_i - \bar{X})^2 = 6(94.56 - 95.87)^2 + \\ &5(99.88 - 95.87)^2 + 5(94.77 - 95.87)^2 + \\ &6(94.75 - 95.87)^2 = 104.27\end{aligned}$$

and the within sample sum of squares

$$SS_w = SS_t - SS_b = 115.63 - 104.27 = 11.36$$

The remainder of the necessary calculations are summarized in the following table.

source	sum-of-squares	degrees of freedom	variance
between samples	104.27	$h - 1 = 4 - 1 = 3$	34.76
within samples	11.36	$N - h = 22 - 4 = 18$	0.631

Comparing the variances we find that

$$F_{\text{exp}} = \frac{s_b^2}{s_w^2} = \frac{34.76}{0.631} = 55.09$$

Because F_{exp} is greater than $F(0.05, 3, 18)$, which is 3.16, we reject the null hypothesis and accept the alternative hypothesis that the work of at least one analyst is significantly different from the remaining analysts. Our best estimate of the within sample variance is

$$\sigma_{\text{rand}}^2 \approx s_w^2 = 0.631$$

and our best estimate of the between sample variance is

$$\sigma_{\text{sys}}^2 \approx \frac{s_b^2 - s_w^2}{n} = \frac{34.76 - 0.631}{22/4} = 6.205$$

In this example the variance due to systematic differences between the analysts is almost an order of magnitude greater than the variance due to the method's precision.

Having demonstrated that there is significant difference between the analysts, we can use a modified version of the t -test—known as **FISHER'S LEAST SIGNIFICANT DIFFERENCE**—to determine the source of the difference. The test statistic for comparing two mean values is the t -test given in equation 4.21 in Chapter 4, except we replace the pooled standard deviation, s_{pool} , by the square root of the within-sample variance from the analysis of variance.

$$t_{\text{exp}} = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{s_w^2}} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad 14.27$$

We compare t_{exp} to its critical value $t(\alpha, \nu)$ using the same significance level as the ANOVA calculation. The degrees of freedom are the same as that for the within sample variance. Since we are interested in whether the larger of the two means is significantly greater than the other mean, the value of $t(\alpha, \nu)$ is that for a one-tailed significance test.

Example 14.9

In [Example 14.8](#) we showed that there is a significant difference between the work of the four analysts in [Table 14.6](#). Determine the source of this significant difference.

SOLUTION

Individual comparisons using Fisher's least significant difference test are based on the following null hypothesis and the appropriate one-tailed alternative hypothesis.

$$H_0: \bar{X}_i = \bar{X}_j \quad H_A: \bar{X}_i > \bar{X}_j \quad \text{or} \quad H_A: \bar{X}_i < \bar{X}_j$$

Using equation 14.27 we calculate values of t_{exp} for each possible comparison and compare them to the one-tailed critical value of 1.73 for $t(0.05, 18)$. For example, t_{exp} for analysts A and B is

$$(t_{\text{exp}})_{AB} = \frac{|94.56 - 99.88|}{\sqrt{0.631}} \times \sqrt{\frac{6 \times 5}{6 + 5}} = 11.06$$

Because $(t_{\text{exp}})_{AB}$ is greater than $t(0.05, 18)$ we reject the null hypothesis and accept the alternative hypothesis that the results for analyst B are significantly greater than those for analyst A. Continuing with the other pairs it is easy to show that $(t_{\text{exp}})_{AC}$ is 0.437, $(t_{\text{exp}})_{AD}$ is 0.414, $(t_{\text{exp}})_{BC}$ is 10.17, $(t_{\text{exp}})_{BD}$ is 10.67, and $(t_{\text{exp}})_{CD}$ is 0.04. Collectively, these results suggest that there is a significant systematic difference between the work of analyst B and the work of the other analysts. There is, of course no way to decide whether any of the four analysts has done accurate work.

You might ask why we bother with the analysis of variance if we are planning to use a t -test to compare pairs of analysts. Each t -test carries a probability, α , of claiming that a difference is significant even though it is not (a type 1 error). If we set α to 0.05 and complete six t -tests, the probability of a type 1 error increases to 0.265. Knowing that there is a significant difference within a data set—what we gain from the analysis of variance—protects the t -test.

We have evidence that analyst B's result is significantly different than the results for analysts A, C, and D, and that we have no evidence that there is any significant difference between the results of analysts A, C, and D. We do not know if analyst B's results are accurate, or if the results of analysts A, C, and D are accurate. In fact, it is possible that none of the results in [Table 14.6](#) are accurate.

We can extend an analysis of variance to systems that involve more than a single variable. For example, we can use a two-way ANOVA to determine the effect on an analytical method of both the analyst and the instrumentation. The treatment of multivariate ANOVA is beyond the scope of this text, but is covered in several of the texts listed in this chapter's additional resources.

14C.3 What is a Reasonable Result for a Collaborative Study?

Collaborative testing provides us with a method for estimating the variability (or reproducibility) between analysts in different labs. If the variability is significant, we can determine what portion is due to indeterminate method errors, σ_{rand}^2 , and what portion is due to systematic differences between the analysts, σ_{syst}^2 . What is left unanswered is the following important question: What is a reasonable value for a method's reproducibility?

An analysis of nearly 10 000 collaborative studies suggests that a reasonable estimate for a method's reproducibility is

$$R = 2^{(1-0.5\log C)} \quad 14.28$$

where R is the percent relative standard deviation for the results included in the collaborative study and C is the fractional amount of analyte in the sample on a weight-to-weight basis.¹⁰ Equation 14.28 is thought to be independent of the type of analyte, the type of matrix, and the method of analysis. For example, when a sample in a collaborative study contains 1 microgram of analyte per gram of sample, C is 10^{-6} and the estimated relative standard deviation is

$$R = 2^{(1-0.5\log 10^{-6})} = 16\%$$

For a discussion of the limitations of equation 14.28, see Linsinger, T. P. J.; Josephs, R. D. "Limitations of the Application of the Horwitz Equation," *Trends Anal. Chem.* **2006**, *25*, 1125–1130, as well as a rebuttal (Thompson, M. "Limitations of the Application of the Horwitz Equation: A Rebuttal," *Trends Anal. Chem.* **2007**, *26*, 659–661) and response to the rebuttal (Linsinger, T. P. J.; Josephs, R. D. "Reply to Professor Michael Thompson's Rebuttal," *Trends Anal. Chem.* **2007**, *26*, 662–663.

Example 14.10

What is the estimated relative standard deviation for the results of a collaborative study when the sample is pure analyte (100% w/w analyte)? Repeat for the case where the analyte's concentration is 0.1% w/w.

SOLUTION

When the sample is 100% w/w analyte ($C=1$) the estimated relative standard deviation is

$$R = 2^{(1-0.5\log 1)} = 2\%$$

We expect that approximately two-thirds of the participants in the collaborative study ($\pm 1\sigma$) will report the analyte's concentration within the range of 98% w/w to 102% w/w. If the analyte's concentration is 0.1% w/w ($C=0.001$), the estimated relative standard deviation is

For a normal distribution, 68.26% of the results fall within $\pm 1\sigma$ of the population's mean (see Table 4.12).

¹⁰ (a) Horwitz, W. *Anal. Chem.* **1982**, *54*, 67A–76A; (b) Hall, P.; Selinger, B. *Anal. Chem.* **1989**, *61*, 1465–1466; (c) Albert, R.; Horwitz, W. *Anal. Chem.* **1997**, *69*, 789–790, (d) "The Amazing Horwitz Function," *AMC Technical Brief 17, July 2004*; (e) Linsinger, T. P. J. *Trends Anal. Chem.* **2006**, *25*, 1125

$$R = 2^{(1-0.5\log 0.001)} = 5.7\%$$

and we expect that approximately two-thirds of the analysts will report the analyte's concentration within the range of 0.094% w/w to 0.106% w/w.

Of course, [equation 14.28](#) only estimates the expected relative standard. If the method's relative standard deviation falls with a range of one-half to twice the estimated value, then it is acceptable for use by analysts in different laboratories. The percent relative standard deviation for a single analyst should be one-half to two-thirds of that for the variability between analysts.

14D Using Excel and R for an Analysis of Variance

Although the calculations for an analysis of variance are relatively straightforward, they become tedious when working with large data sets. Both Excel and R include functions for completing an analysis of variance. In addition, R provides a function for identifying the source(s) of significant differences within the data set.

14D.1 Excel

Excel's Analysis ToolPak includes a tool to help you complete an analysis of variance. Let's use the ToolPak to complete an analysis of variance on the data in [Table 14.6](#). Enter the data from [Table 14.6](#) into a spreadsheet as shown in [Figure 14.22](#). To complete the analysis of variance select **Data Analysis...** from the **Tools** menu, which opens a window entitled "*Data Analysis*." Scroll through the window, select **Analysis: Single Factor** from the available options and click **OK**. Place the cursor in the box for the "*Input range*" and then click and drag over the cells B1:E7. Select the radio button for "*Grouped by: columns*" and check the box for "*Labels in the first row*." In the box for "*Alpha*" enter 0.05 for α . Select the radio button for "*Output range*," place the cursor in the box and click on an empty cell; this is where Excel will place the results. Clicking **OK** generates the information shown in [Figure 14.23](#). The small value of 3.05×10^{-9} for falsely rejecting the null hypothesis indicates that there is a significant source of variation between the analysts.

Excel's Data Analysis Toolpak is available for Windows. Older versions of Excel for Mac include the toolpak; however, beginning with Excel for Mac 2011, the toolpak no longer is available.

	A	B	C	D	E
1	replicate	analyst A	analyst B	analyst C	analyst D
2	1	94.09	99.55	95.14	93.88
3	2	94.64	98.24	94.62	94.23
4	3	95.08	101.1	95.28	96.05
5	4	94.54	100.4	94.59	93.89
6	5	95.38	100.1	94.24	94.59
7	6	93.62			95.49

Figure 14.22 Portion of a spreadsheet containing the data from [Table 14.6](#).

Anova: Single Factor					
SUMMARY					
Groups	Count	Sum	Average	Variance	
analyst A	6	567.35	94.5583333	0.41081667	
analyst B	5	499.39	99.878	1.15142	
analyst C	5	473.87	94.774	0.18318	
analyst D	6	568.49	94.7483333	0.80889667	

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	104.197961	3	34.7326535	54.6637742	3.0463E-09	3.1599076
Within Groups	11.4369667	18	0.63538704			
Total	115.634927	21				

Figure 14.23 Output from Excel's one-way analysis of variance of the data in [Table 14.6](#). The summary table provides the mean and variance for each analyst. The ANOVA table summarizes the sum-of-squares terms (*SS*), the degrees of freedom (*df*), the variances (*MS* for mean square), the value of F_{exp} and the critical value of *F*, and the probability of incorrectly rejecting the null hypothesis that there is no significant difference between the analysts.

14D.2 R

To complete an analysis of variance for the data in [Table 14.6](#) using R, we first need to create several objects. The first object contains each result from [Table 14.6](#).

```
> results=c(94.090, 94.640, 95.008, 94.540, 95.380, 93.620,
99.550, 98.240, 101.100, 100.400, 100.100, 95.140, 94.620,
95.280, 94.590, 94.240, 93.880, 94.230, 96.050, 93.890,
94.950, 95.490)
```

The second object contains labels that identify the source of each entry in the first object. The following code creates this object.

```
> analyst=c(rep("a",6), rep("b",5), rep("c",5), rep("d",6))
```

Next, we combine the two objects into a table with two columns, one that contains the data (*results*) and one that contains the labels (*analyst*).

```
> df=data.frame(results, labels=factor(analyst))
```

The command **factor** indicates that the object *analyst* contains the categorical factors for the analysis of variance. The command for an analysis of variance takes the following form

```
anova(lm(data ~ factors), data = data.frame)
```

where *data* and *factors* are the columns that contain the data and the categorical factors, and *data.frame* is the name we assigned to the data table. [Figure 14.24](#) shows the output for an analysis of variance of the data in [Table 14.6](#). The small value of 3.04×10^{-9} for falsely rejecting the null hypothesis indicates that there is a significant source of variation between the analysts.

Having found a significant difference between the analysts, we want to identify the source of this difference. R does not include Fisher's least sig-

You can arrange the results in any order. In creating this object, I choose to list the results for analyst *A*, followed by the results for analyst *B*, *C*, and *D*.

The command **rep** (for repeat) has two variables: the item to repeat and the number of times it is repeated. The object *analyst* is the vector ("a", "a", "a", "a", "a", "a", "b", "b", "b", "b", "b", "c", "c", "c", "c", "c", "d", "d", "d", "d", "d", "d").

We call this table a data frame. Many functions in R work on the columns in a data frame.

The command **lm** stands for linear model. See Section 5F.2 in Chapter 5 for a discussion of linear models in R.

```
> anova(lm(results ~ labels, data = df))

Analysis of Variance Table

Response: results
      Df Sum Sq Mean Sq F value Pr(>F)
labels  3  104.198   34.733   54.664 3.04e-09 ***
Residuals 18    11.366    0.631
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 14.24 Output of an R session for an analysis of variance for the data in [Table 14.6](#). In the table, “labels” is the between-sample variance and “residuals” is the within-sample variance. The p -value of $3.04e-09$ is the probability of incorrectly rejecting the null hypothesis that the within-sample and between-sample variances are the same.

nificant difference test, but it does include a function for a related method called Tukey’s honest significant difference test. The command for this test takes the following form

```
> TukeyHSD(aov(lm(data ~ factors), data = data.frame), conf,
level=0.95)
```

where *data* and *factors* are the columns that contain the data and the categorical factors, and *data.frame* is the name we assigned to the data table. Figure 14.25 shows the output of this command and its interpretation. The small probability values when comparing analyst B to each of the other analysts indicates that this is the source of the significant difference identified in the analysis of variance.

```
> TukeyHSD(aov(results ~ labels, data = df))
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = results ~ labels, data = df)

$labels
      diff      lwr      upr      p adj
b-a  5.31966667  3.928277  6.711057  0.0000000
c-a  0.21566667 -1.175723  1.607057  0.9710635
d-a  0.28000000 -1.046638  1.606638  0.9318110
c-b -5.10400000 -6.557260 -3.650740  0.0000001
d-b -5.03966667 -6.431057 -3.648277  0.0000000
d-c  0.06433333 -1.327057  1.455723  0.9991718
```

Figure 14.25 Output of an R session for a Tukey honest significance difference test using the data in [Table 14.6](#). For each possible comparison of analysts, the table gives the actual difference between the analysts, “diff,” and the smallest, “lwr,” and the largest, “upr,” differences for a 95% confidence interval. The “p adj” is the probability that a difference of zero falls within this confidence interval. The smaller the p -value, the greater the probability that the difference between the analysts is significant.

You may recall that an underlined command is the default value. If you are using an α of 0.05 (a 95% confidence level), then you do not need to include the entry for `conf.level`. If you wish to use an α of 0.10, then enter `conf.level=0.90`.

Note that p value is small when the confidence interval for the difference includes zero.

14E Key Terms

2^k factorial design	analysis of variance	between-sample variance
blind analysis	central composite design	collaborative testing
dependent	effective	efficiency
empirical model	factor	factor level
Fisher's least significant difference	fixed-size simplex optimization	global optimum
independent	local optimum	one-factor-at-a-time optimization
response	response surface	ruggedness testing
searching algorithm	simplex	standard method
theoretical model	validation	variable-sized simplex optimization
within-sample variance		

14F Summary

One of the goals of analytical chemistry is to develop new analytical methods that are accepted as standard methods. In this chapter we have considered how a standard method is developed, including finding the optimum experimental conditions, verifying that the method produces acceptable precision and accuracy, and validating the method for general use.

To optimize a method we try to find the combination of experimental parameters that produces the best result or response. We can visualize this process as being similar to finding the highest point on a mountain. In this analogy, the mountain's topography corresponds to a response surface, which is a plot of the system's response as a function of the factors under our control.

One method for finding the optimum response is to use a searching algorithm. In a one-factor-at-a-time optimization, we change one factor while holding constant all other factors until there is no further improvement in the response. The process continues with the next factor, cycling through the factors until there is no further improvement in the response. This approach to finding the optimum response often is effective, but usually is not efficient. A searching algorithm that is both effective and efficient is a simplex optimization, the rules of which allow us to change the levels of all factors simultaneously.

Another approach to optimizing a method is to develop a mathematical model of the response surface. Such models can be theoretical, in that they are derived from a known chemical and physical relationship between the response and its factors. Alternatively, we can develop an empirical model, which does not have a firm theoretical basis, by fitting an empirical equation to our experimental data. One approach is to use a 2^k factorial design

in which each factor is tested at both a high level and a low level, and paired with the high level and the low level for all other factors.

After optimizing a method it is necessary to demonstrate that it can produce acceptable results. Verifying a method usually includes establishing single-operator characteristics, the blind analysis of standard samples, and determining the method's ruggedness. Single-operator characteristics include the method's precision, accuracy, and detection limit when used by a single analyst. To test against possible bias on the part of the analyst, he or she analyzes a set of blind samples in which the analyst does not know the concentration of analyte. Finally, we use ruggedness testing to determine which experimental factors must be carefully controlled to avoid unexpectedly large determinate or indeterminate sources of error.

The last step in establishing a standard method is to validate its transferability to other laboratories. An important step in the process of validating a method is collaborative testing, in which a common set of samples is analyzed by different laboratories. In a well-designed collaborative test it is possible to establish limits for the method's precision and accuracy.

14G Problems

- For each of the following equations determine the optimum response using a one-factor-at-a-time searching algorithm. Begin the search at (0,0) by first changing factor A , using a step-size of 1 for both factors. The boundary conditions for each response surface are $0 \leq A \leq 10$ and $0 \leq B \leq 10$. Continue the search through as many cycles as necessary until you find the optimum response. Compare your optimum response for each equation to the true optimum.

$$(a) R = 1.68 + 0.24A + 0.56B - 0.04A^2 - 0.04B^2 \quad \mu_{\text{opt}} = (3, 7)$$

$$(b) R = 4.0 - 0.4A + 0.08AB \quad \mu_{\text{opt}} = (10, 10)$$

$$(c) R = 3.264 + 1.537A + 0.5664B - 0.1505A^2 - 0.02734B^2 - 0.05785AB \quad \mu_{\text{opt}} = (391, 6.22)$$

- Use a fixed-sized simplex searching algorithm to find the optimum response for the equation in Problem 1c. For the first simplex, set one vertex at (0,0) with step sizes of one. Compare your optimum response to the true optimum.
- Show that [equation 14.3](#) and [equation 14.4](#) are correct.
- A 2^k factorial design was used to determine the equation for the response surface in Problem 1b. The uncoded levels, coded levels, and the responses are shown in the following table. Determine the uncoded equation for the response surface.

Note: These equations are from Deming, S. N.; Morgan, S. L. *Experimental Design: A Chemometric Approach*, Elsevier: Amsterdam, 1987, and pseudo-three dimensional plots of the response surfaces can be found in their Figures 11.4, 11.5 and 11.14.

<i>A</i>	<i>B</i>	<i>A</i> *	<i>B</i> *	response
8	8	+1	+1	5.92
8	2	+1	-1	2.08
2	8	-1	+1	4.48
2	2	-1	-1	3.52

5. Koscielniak and Parczewski investigated the influence of Al on the determination of Ca by atomic absorption spectrophotometry using the 2^k factorial design shown in the following table.¹¹

Ca ²⁺ (ppm)	Al ³⁺ (ppm)	<i>Ca</i> *	<i>Al</i> *	response
10	160	+1	+1	54.92
10	0	+1	-1	98.44
4	160	-1	+1	19.18
4	0	-1	-1	38.52

- (a) Determine the uncoded equation for the response surface.
- (b) If you wish to analyze a sample that is 6.0 ppm Ca²⁺, what is the maximum concentration of Al³⁺ that can be present if the error in the response must be less than 5.0%?
6. Strange studied a chemical reaction using a 2^3 factorial design.¹²

factor	high (+1) level	low (-1) level
X: temperature	140 °C	120 °C
Y: catalyst	type B	type A
Z: [reactant]	0.50 M	0.25 M

run	<i>X</i> *	<i>Y</i> *	<i>Z</i> *	% yield
1	-1	-1	-1	28
2	+1	-1	-1	17
3	-1	+1	-1	41
4	+1	+1	-1	34
5	-1	-1	+1	56
6	+1	-1	+1	51
7	-1	+1	+1	42
8	+1	+1	+1	36

¹¹ Koscielniak, P.; Parczewski, A. *Anal. Chim. Acta* **1983**, *153*, 111–119.

¹² Strange, R. S. *J. Chem. Educ.* **1990**, *67*, 113–115.

- (a) Determine the coded equation for this data.
- (b) If β terms of less than ± 1 are insignificant, what main effects and what interaction terms in the coded equation are important? Write down this simpler form for the coded equation.
- (c) Explain why the coded equation for this data can not be transformed into an uncoded form.
- (d) Which is the better catalyst, A or B?
- (e) What is the yield if the temperature is set to 125 °C, the concentration of the reactant is 0.45 M, and we use the appropriate catalyst?
7. Pharmaceutical tablets coated with lactose often develop a brown discoloration. The primary factors that affect the discoloration are temperature, relative humidity, and the presence of a base acting as a catalyst. The following data have been reported for a 2^3 factorial design.¹³

factor	high (+1) level	low (-1) level
X: benzocaine	present	absent
Y: temperature	40 °C	25 °C
Z: relative humidity	75%	50%

run	X^*	Y^*	Z^*	color (arb. units)
1	-1	-1	-1	1.55
2	+1	-1	-1	5.40
3	-1	+1	-1	3.50
4	+1	+1	-1	6.75
5	-1	-1	+1	2.45
6	+1	-1	+1	3.60
7	-1	+1	+1	3.05
8	+1	+1	+1	7.10

- (a) Determine the coded equation for this data.
- (b) If β terms of less than 0.5 are insignificant, what main effects and what interaction terms in the coded equation are important? Write down this simpler form for the coded equation.

¹³ Armstrong, N. A.; James, K. C. *Pharmaceutical Experimental Design and Interpretation*, Taylor and Francis: London, 1996 as cited in Gonzalez, A. G. *Anal. Chim. Acta* **1998**, *360*, 227–241.

8. The following data for a 2^3 factorial design were collected during a study of the effect of temperature, pressure, and residence time on the % yield of a reaction.¹⁴

factor	high (+1) level	low (-1) level
X: temperature	200 °C	100 °C
Y: pressure	0.6 MPa	0.2 MPa
Z: residence time	20 min	10 min

run	X^*	Y^*	Z^*	percent yield
1	-1	-1	-1	2
2	+1	-1	-1	6
3	-1	+1	-1	4
4	+1	+1	-1	8
5	-1	-1	+1	10
6	+1	-1	+1	18
7	-1	+1	+1	8
8	+1	+1	+1	12

- (a) Determine the coded equation for this data.
- (b) If β terms of less than 0.5 are insignificant, what main effects and what interaction terms in the coded equation are important? Write down this simpler form for the coded equation.
- (c) Three runs at the center of the factorial design—a temperature of 150 °C, a pressure of 0.4 MPa, and a residence time of 15 min—give percent yields of 8%, 9%, and 8.8%. Determine if a first-order empirical model is appropriate for this system at $\alpha = 0.05$.
9. Duarte and colleagues used a factorial design to optimize a flow-injection analysis method for determining penicillin.¹⁵ Three factors were studied: reactor length, carrier flow rate, and sample volume, with the high and low values summarized in the following table.

factor	high (+1) level	low (-1) level
X: reactor length	1.5 cm	2.0 cm
Y: carrier flow rate	1.6 mL/min	2.2 mL/min
Z: sample volume	100 μ L	150 μ L

Note that the coded values +1 and -1 need not correspond to physical larger and physically smaller values. In this case, for example, all three factors have their largest value assigned to the low, or -1 level.

¹⁴ Akhnazarova, S.; Kafarov, V. *Experimental Optimization in Chemistry and Chemical Engineering*, MIR Publishers: Moscow, 1982 as cited in Gonzalez, A. G. *Anal. Chim. Acta* **1998**, *360*, 227–241.

¹⁵ Duarte, M. M. M. B.; de O. Netro, G.; Kubota, L. T.; Filho, J. L. L.; Pimentel, M. F.; Lima, F.; Lins, V. *Anal. Chim. Acta* **1997**, *350*, 353–357.

The authors determined the optimum response using two criteria: the greatest sensitivity, as determined by the change in potential for the potentiometric detector, and the largest sampling rate. The following table summarizes their optimization results.

run	X^*	Y^*	Z^*	ΔE (mV)	samples/h
1	-1	-1	-1	37.45	21.5
2	+1	-1	-1	31.70	26.0
3	-1	+1	-1	32.10	30.0
4	+1	+1	-1	27.20	33.0
5	-1	-1	+1	39.85	21.0
6	+1	-1	+1	32.85	19.5
7	-1	+1	+1	35.00	30.0
8	+1	+1	+1	32.15	34.0

- Determine the coded equation for the response surface where ΔE is the response.
 - Determine the coded equation for the response surface where sample/h is the response.
 - Based on the coded equations in (a) and in (b), do conditions that favor sensitivity also improve the sampling rate?
 - What conditions would you choose if your goal is to optimize both sensitivity and sampling rate?
10. Here is a challenge! McMinn, Eatherton, and Hill investigated the effect of five factors for optimizing an H_2 -atmosphere flame ionization detector using a 2^5 factorial design.¹⁶ The factors and their levels were

factor	high (+1) level	low (-1) level
A: H_2 flow rate	1460 mL/min	1382 mL/min
B: SiH_4	20.0 ppm	12.2 ppm
C: $O_2 + N_2$ flow rate	255 mL/min	210 mL/min
D: O_2/N_2	1.36	1.19
E: electrode height	75 (arb. unit)	55 (arb. unit)

The coded (“+” = +1, “-” = -1) factor levels and responses, R , for the 32 experiments are shown in the following table

¹⁶ McMinn, D. G.; Eatherton, R. L.; Hill, H. H. *Anal. Chem.* **1984**, *56*, 1293–1298.

run	A^*	B^*	C^*	D^*	E^*	run	A^*	B^*	C^*	D^*	E^*
1	-	-	-	-	-	17	-	-	-	-	+
2	+	-	-	-	-	18	+	-	-	-	+
3	-	+	-	-	-	19	-	+	-	-	+
4	+	+	-	-	-	20	+	+	-	-	+
5	-	-	+	-	-	21	-	-	+	-	+
6	+	-	+	-	-	22	+	-	+	-	+
7	-	+	+	-	-	23	-	+	+	-	+
8	+	+	+	-	-	24	+	+	+	-	+
9	-	-	-	+	-	25	-	-	-	+	+
10	+	-	-	+	-	26	+	-	-	+	+
11	-	+	-	+	-	27	-	+	-	+	+
12	+	+	-	+	-	28	+	+	-	+	+
13	-	+	+	+	-	29	-	-	+	+	+
14	+	-	+	+	-	30	+	-	+	+	+
15	-	+	+	+	-	31	-	+	+	+	+
16	+	+	+	+	-	32	+	+	+	+	+

- (a) Determine the coded equation for this response surface, ignoring β terms less than ± 0.03 .
- (b) A simplex optimization of this system finds optimal values for the factors of $A=2278$ mL/min, $B=9.90$ ppm, $C=260.6$ mL/min, and $D=1.71$. The value of E was maintained at its high level. Are these values consistent with your analysis of the factorial design.

11. A good empirical model provides an accurate picture of the response surface over the range of factor levels within the experimental design. The same model, however, may yield an inaccurate prediction for the response at other factor levels. For this reason, an empirical model, is tested before it is extrapolated to conditions other than those used in determining the model. For example, Palasota and Deming studied the effect of the relative amounts of H_2SO_4 and H_2O_2 on the absorbance of solutions of vanadium using the following central composite design.¹⁷

run	drops 1% H_2SO_4	drops 20% H_2O_2
1	15	22
2	10	20
3	20	20

¹⁷ Palasota, J. A.; Deming, S. N. *J. Chem. Educ.* **1992**, *62*, 560–563.

run	drops 1% H ₂ SO ₄	drops 20% H ₂ O ₂
4	8	15
5	15	15
6	15	15
7	15	15
8	15	15
9	22	15
10	10	10
11	20	10
12	15	8

The reaction of H₂SO₄ and H₂O₂ generates a red-brown solution whose absorbance is measured at a wavelength of 450 nm. A regression analysis on their data yields the following uncoded equation for the response (absorbance × 1000).

$$R = 835.90 - 36.82X_1 - 21.34X_2 + 0.52X_1^2 + 0.15X_2^2 + 0.98X_1X_2$$

where X_1 is the drops of H₂O₂, and X_2 is the drops of H₂SO₄. Calculate the predicted absorbances for 10 drops of H₂O₂ and 0 drops of H₂SO₄, 0 drops of H₂O₂ and 10 drops of H₂SO₄, and for 0 drops of each reagent. Are these results reasonable? Explain. What does your answer tell you about this empirical model?

12. A newly proposed method is tested for its single-operator characteristics. To be competitive with the standard method, the new method must have a relative standard deviation of less than 10%, with a bias of less than 10%. To test the method, an analyst performs 10 replicate analyses on a standard sample known to contain 1.30 ppm of analyte. The results for the 10 trials are

1.25 1.26 1.29 1.56 1.46 1.23 1.49 1.27 1.31 1.43

Are the single-operator characteristics for this method acceptable?

13. A proposed gravimetric method was evaluated for its ruggedness by varying the following factors.

Factor A: sample size	$A = 1$ g	$a = 1.1$ g
Factor B: pH	$B = 6.5$	$b = 6.0$
Factor C: digestion time	$C = 3$ h	$c = 1$ h
Factor D: number rinses	$D = 3$	$d = 5$

Factor E: precipitant	$E =$ reagent 1	$e =$ reagent 2
Factor F: digestion temperature	$F = 50^\circ\text{C}$	$f = 60^\circ\text{C}$
Factor G: drying temperature	$G = 110^\circ\text{C}$	$g = 140^\circ\text{C}$

A standard sample that contains a known amount of analyte is carried through the procedure using the experimental design in [Table 14.5](#). The percentage of analyte actually found in the eight trials are as follows: $R_1 = 98.9$, $R_2 = 98.5$, $R_3 = 97.7$, $R_4 = 97.0$, $R_5 = 98.8$, $R_6 = 98.5$, $R_7 = 97.7$, and $R_8 = 97.3$. Determine which factors, if any, appear to have a significant affect on the response, and estimate the expected standard deviation for the method.

14. The two-sample plot for the data in [Example 14.6](#) is shown in [Figure 14.21](#). Identify the analyst whose work is (a) the most accurate, (b) the most precise, (c) the least accurate, and (d) the least precise.
15. Chichilo reports the following data for the determination of the %w/w Al in two samples of limestone.¹⁸

analyst	sample 1	sample 2
1	1.35	1.57
2	1.35	1.33
3	1.34	1.47
4	1.50	1.60
5	1.52	1.62
6	1.39	1.52
7	1.30	1.36
8	1.32	1.53

Construct a two-sample plot for this data and estimate values for σ_{rand} and for σ_{syst} .

16. The importance of between-laboratory variability on the results of an analytical method are determined by having several laboratories analyze the same sample. In one such study, seven laboratories analyzed a sample of homogenized milk for a selected aflatoxin.¹⁹ The results, in ppb, are summarized below.

¹⁸ Chichilo, P. J. *J. Assoc. Offic. Agr. Chemists* **1964**, *47*, 1019 as reported in Youden, W. J. "Statistical Techniques for Collaborative Tests," in *Statistical Manual of the Association of Official Analytical Chemists*, Association of Official Analytical Chemists: Washington, D. C., 1975

¹⁹ Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Chemometrics: A Textbook*, Elsevier: Amsterdam, 1988.

lab A	lab B	lab C	lab D	lab E	lab F	lab G
1.6	4.6	1.2	1.5	6.0	6.2	3.3
2.9	2.8	1.9	2.7	3.9	3.8	3.8
3.5	3.0	2.9	3.4	4.3	5.5	5.5
1.8	4.5	1.1	2.0	5.8	4.2	4.9
2.2	3.1	2.9	3.4	4.0	5.3	4.5

- (a) Determine if the between-laboratory variability is significantly greater than the within-laboratory variability at $\alpha = 0.05$. If the between-laboratory variability is significant, then determine the source(s) of that variability.
- (b) Estimate values for σ_{rand}^2 and for σ_{sys}^2 .
17. Show that the total sum-of-squares (SS_t) is the sum of the within-sample sum-of-squares (SS_w) and the between-sample sum-of-squares (SS_b). See [Table 14.7](#) for the relevant equations.
18. Eighteen analytical students are asked to determine the %w/w Mn in a sample of steel, with the results shown here.

0.26% 0.28% 0.27% 0.24% 0.26% 0.25%

0.26% 0.28% 0.25% 0.24% 0.26% 0.25%

0.29% 0.24% 0.27% 0.23% 0.26% 0.24%

- (a) Given that the steel sample is 0.26% w/w Mn, estimate the expected relative standard deviation for the class' results.
- (b) The actual results obtained by the students are shown here. Are these results consistent with the estimated relative standard deviation?

14H Solutions to Practice Exercises

Practice Exercise 14.1

If we hold factor A at level A_1 , changing factor B from level B_1 to level B_2 increases the response from 40 to 60, or a change ΔR , of

$$\Delta R = 60 - 40 = 20$$

If we hold factor A at level A_2 , we find that we have the same change in response when the level of factor B changes from B_1 to B_2 .

$$\Delta R = 100 - 80 = 20$$

[Click here](#) to return to the chapter.

Practice Exercise 14.2

If we hold factor B at level B_1 , changing factor A from level A_1 to level A_2 increases the response from 20 to 80, or a change ΔR , of

$$\Delta R = 80 - 20 = 60$$

If we hold factor B at level B_2 , we find that the change in response when the level of factor A changes from A_1 to A_2 is now 20.

$$\Delta R = 80 - 60 = 20$$

Click [here](#) to return to the chapter.
