10 different methods of teaching a foreign language in elementary school and that we had no particular reason to expect any one method to be better than any other. How might we analyze the results of this experiment? One procedure would be to treat each of the possible two-group comparisons as a different *two-group experiment*. That is, we would compare method 1 versus methods 2, 3, . . . , 9, and 10; method 2 versus methods 3, 4, . . . , 9, and 10; and so on. There are 45 of these two-group comparisons. Obviously, this sort of analysis would require a considerable amount of calculation. Moreover, we should be concerned with the fact that we are using the same sets of data over and over again to make these comparisons. (Actually, we are using each group a total of 9 times.) We cannot think of these comparisons as constituting 45 *independent* experiments; if one group is distorted for some reason or other, this distortion will be present in all 9 of the comparisons in which it enters.

The single-factor analysis of variance allows us to consider all of the treatments in a *single* assessment. Without going into the details, this analysis sets in perspective any interpretations we may want to make concerning the differences we have observed. More specifically, the analysis will tell us whether or not it will be worthwhile to conduct any additional analyses comparing specific treatment groups.

I will first consider the logic behind the analysis of variance and then worry about translating these intuitive notions into mathematical expressions and actual numbers.

2

Specifying Sources of Variability

As described in Chap. 1, the first step in testing a research hypothesis is to design an experiment in which the influence of known and unknown variables is minimized. If uncontrolled, such variables could result in a systematic bias, that is, a confounding with the independent variable. Although it is possible to control nuisance variables by holding them constant, the most common procedure is to spread their effects randomly over all the treatment conditions. Unfortunately, however, the use of randomization to control the influence of nuisance variables—which is nearly unavoidable in experimentation—creates a new problem:

> **Differences observed among treatment means are influenced jointly by the actual differences in the treatments administered to the different groups *and* by chance factors introduced by randomization.**

The decision confronting the experimenter is to decide whether the differences associated with the treatment conditions are entirely or just partly due to chance. I will now consider, in general terms, a statistical solution to this disturbing problem.

2.1 THE LOGIC OF HYPOTHESIS TESTING

Suppose we have just completed collecting the data from an experiment. As a first step in the analysis of the data, we calculate summary statistics for each treatment condition—usually measures of "average" performance (the mean) and variability (the variance or standard deviation). Generally, we are not primarily interested merely in describing the performance of subjects in the different treatment conditions. Our main goal is to make *inferences* about the behavior of subjects who have not been tested in our experiment. Rarely will we choose to test all possible subjects in an experiment, such as all laboratory rats of a particular strain or all college students enrolled in an introductory psychology class at a particular university. Instead, we select samples from these larger groups, administer the experimental conditions to the samples, and make inferences about the nature of the population on the basis of the experimental outcome. We refer to these large groups as **populations**. Members of any population are identified by a set of rules of membership. A **sample** consists of a smaller set of observations drawn from the population. To be able to generalize to the population in a strict statistical sense, we must select the subjects constituting the sample *randomly* from the population. Summary descriptions calculated from the data of a sample are called **statistics**, and measures calculated from all the observations within the population are called **parameters**. In most cases, I will use Roman letters to designate statistics and Greek letters to designate parameters.

At this point, we can view the subjects in the treatment conditions as representing samples drawn from different treatment populations. Statistics, calculated on the scores obtained from the different groups of subjects, provide estimates of one or more parameters for the different treatment populations. We are now ready to consider the formal process of hypothesis testing, where we translate the research hypothesis into a set of **statistical hypotheses**, which we then evaluate in light of the obtained data.

Statistical Hypotheses

A research hypothesis is a fairly general statement about the assumed nature of the world that we translate into an experiment. Typically, but not always, a research hypothesis asserts that the treatments will produce an effect. (If it did not, we would probably not have performed the experiment in the first place!) Statistical hypotheses consist of a set of precise hypotheses about the parameters of the different treatment populations. We usually formulate two statistical hypotheses that are mutually exclusive or incompatible statements about the treatment parameters.

The statistical hypothesis we *test* is called the **null hypothesis**, often symbolized as H_0. The function of the null hypothesis is to specify the values of a particular parameter (the mean, for example) in the different treatment populations (symbolized as μ_1, μ_2, μ_3, and so on). The null hypothesis typically chosen gives the *same* value to the different populations—that is,

$$H_0: \mu_1 = \mu_2 = \mu_3 = \text{etc.}$$

This is tantamount to saying that *no* treatment effects are present in the population. If the actual means obtained from the treatment groups are too deviant from those specified by the null hypothesis, H_0 is rejected in favor of the other statistical hypothesis, called the **alternative hypothesis**, H_1. The alternative hypothesis specifies values for the parameter that are *incompatible* with the null hypothesis. Usually, the alternative hypothesis states simply that the values of the parameter in the different treatment populations are *not* all equal. Specifically,

$$H_1: \text{not all } \mu\text{'s are equal.}$$

A decision to reject H_0 implies an acceptance of H_1, which, in essence, constitutes support of our original *research* hypothesis. On the other hand, if the treatment means are reasonably close to those specified by the null hypothesis, H_0 is retained and not rejected. This latter decision can be thought of as a failure of the experiment to support the research hypothesis. You will see in a later discussion that a decision to retain the null hypothesis is not simple. Depending on the true state of the world, that is, the equality or inequality of the actual treatment population means, we can make an error of inference with *either* decision, rejecting H_0 or retaining H_0. (I will say more about these errors in Chap. 3.)

Experimental Error

The crux of the problem is the fact that we can always attribute some portion of the differences we observe among the treatment means to chance factors. All nuisance variables that we control in our experiment through random assignment of subjects to the treatment conditions are considered potential contributors to **exper-**

imental error. In the behavioral sciences, the most important source of experimental error is that due to individual differences. In Chap. 1, I also mentioned variations in the various features of the testing environment as contributing to uncontrolled variability. Another source of experimental error is what may be called *measurement error*. A misreading of a dial, a misjudgment that a particular type of behavior has occurred, the variability in reaction time of an experimenter timing a given bit of behavior, and an error in transposing observations recorded in the laboratory to summary worksheets used in performing the statistical analyses are all examples of measurement error. Although it is not obvious, a given experimental treatment is not exactly the same for each subject serving in that treatment condition; the experimental apparatus cannot be counted on to administer the same treatment for successive subjects. An experimenter cannot construct an identical testing environment (the reading of instructions, the experimenter-subject interaction, and so on) for all subjects in any treatment group. We describe the contribution of all these different components of experimental error as *unsystematic*, stressing the fact that their influence is *independent* of the treatment effects.

Estimates of Experimental Error

Suppose we were able to estimate the extent to which the differences we observe among the group means are due to experimental error. We would then be in a position to begin to consider the evaluation of the hypothesis that the means of the treatment populations are equal. Consider the scores of subjects in any one of the treatment conditions. We certainly do not expect these scores to be equal. In the ideal experiment they would be. In an actual experiment, of course, all the sources of uncontrolled variability will also contribute to a subject's score, resulting in a difference in performance for subjects who are given the same treatment condition. The variability of subjects treated alike, that is, within the same treatment level, provides an estimate of experimental error. By the same argument, the variability of subjects within each of the other treatment levels also offers estimates of experimental error. If we assume that experimental error is the same for the different treatment conditions, we can obtain a more stable estimate of this quantity by pooling and averaging these separate estimates.

Assume that we have drawn random samples from a population of subjects, administered the different treatments, recorded the performance of the subjects, and calculated the means of the treatment groups. Further assume for the moment that the null hypothesis is *true*—that the population means associated with the treatment conditions are *equal*. Would we expect our *sample* means, the means calculated in the experiment, to be equal? Certainly not. From our discussion of the use of randomization to "control" unwanted factors in our experiment, it should be clear that the means will rarely be equal. If the sample means are not equal, the only reasonable explanation that we can offer for these differences is the operation of experimental error. All the sources of unsystematic variability, which contribute to the differences among subjects within a given treatment condition, will also be operating to produce differences among the sample means.

Take, for instance, error that results from the random assignment of subjects to treatments. If the assignment procedure is truly random, each subject will have an equal chance of being assigned to any one of the different treatments. But this in no way *guarantees* that the average ability of subjects assigned to these groups is equal. Similarly, for the other contributors to experimental error, there is no reason to expect these uncontrolled sources of error to balance out perfectly across the treatment conditions. In short, then, under these circumstances—an experiment conducted when the null hypothesis is true—differences among the sample means will reflect the operation of experimental error.

Estimates of Treatment Effects

So far in this discussion I have considered only the case in which the null hypothesis is true. Certainly we hope that we will discover at least a few situations in which the null hypothesis is *false*, in which case, there are real differences among the means of the treatment populations. Assuming that the subjects in each treatment group are drawn randomly from corresponding treatment populations, the means of the different groups in the experiment should reflect only the differences in the population means. The mere fact that the null hypothesis is false does not imply that experimental error has vanished, however. Not at all. The only change is that there is now an additional component contributing to the differences among the means, a systematic component as opposed to an unsystematic one, namely, **treatment effects**.

Thus, differences among treatment means may reflect *two different quantities*: When the population means are equal, the differences among the group means will reflect the operation of experimental error alone, but when the population means are not equal, the differences among the group means will reflect the operation of an unsystematic component and a systematic component, that is, experimental error and treatment effects, respectively.

Evaluation of the Null Hypothesis

You have seen that when the null hypothesis is true, we will have two estimates of experimental error available from the experiment. If we form a *ratio* of these two estimates, we will find that we have produced a useful statistic. More specifically, consider the following ratio:

$$\frac{\text{differences among treatment means}}{\text{differences among subjects treated alike}}$$

From this discussion, we can think of this ratio as contrasting an estimate of experimental error based on differences between groups with an estimate of experimental error based on pooled differences within groups. That is, we have

$$\frac{\text{experimental error}}{\text{experimental error}}$$

If we were to repeat this experiment a large number of times on new samples of subjects drawn from the same population, we would expect to find an average value of this ratio of approximately 1.0.

Consider now the same ratio when the null hypothesis is *false*. Under these circumstances, there is an additional component in the numerator, one that reflects the treatment effects. Explicitly, the ratio becomes

$$\frac{\text{(treatment effects)} + \text{(experimental error)}}{\text{experimental error}}$$

Given this situation, if we were to repeat the experiment a large number of times, we would expect to find an average value of this ratio that is *greater* than 1.0.

You can see, then, that the average value of this ratio, obtained from a large number of replications of the experiment, depends on the values of the population means. If H_0 is true (that is, the means are equal), the average value will approximate 1.0; if H_1 is true (that is, the means are not equal), the average value will approximate a number greater than 1.0. A problem remains, however, since in any one experiment, it is always possible to obtain a value that is *greater* than 1.0 when H_0 is *true* and one that is *equal* to or *less* than 1.0 when H_1 is *true*. Thus, merely checking to see whether the ratio is greater than 1.0 does not tell us which statistical hypothesis is correct.

What we will do about this problem is to make a decision concerning the acceptability of the null hypothesis that is based on a consideration of the chance probability associated with the ratio we actually found in the experiment. If the probability of obtaining a ratio of this size or larger by chance is reasonably low, we will reject the null hypothesis. On the other hand, if this probability is high, we will not reject, or in essence, we will retain the null hypothesis. (I will have more to say about the **decision rules** we follow in making this decision in Chap. 3.)

2.2 THE COMPONENT DEVIATIONS

In the remainder of this chapter, you will see the abstract notions of variability *between* treatment groups and *within* treatment groups become concrete arithmetic operations extracted from scores produced in single-factor experiments. The next chapter indicates how we use this information to provide a test of the null hypothesis.

Suppose we were interested in the effect on reading comprehension of three different instructions. One group of children is asked to attempt to memorize an essay, a second group is asked to concentrate on the ideas in the essay, and a third group is given no specific instructions. I will refer to the independent variable, types of instruction, as **factor** A, and to the three levels of factor A (the three different instructional conditions) as levels a_1, a_2, and a_3. We draw the subjects from a fourth-grade class and randomly assign $n = 5$ different subjects to each of the levels of factor A. We allow all subjects to study the essay for 10 minutes, after which time we give them an objective test to determine their comprehension of the

Table 2–1 Numerical Example

	FACTOR A	
Level a_1	**Level a_2**	**Level a_3**
16	4	2
18	6	10
10	8	9
12	10	13
19	2	11

passage. The response measure—that is, the score for each subject (Y)—is the number of test items correctly answered by each student.

The data from this hypothetical experiment are presented in Table 2–1. The five Y scores for each treatment condition are arranged in three columns. Our first step in the analysis would be to compute the means for the three sets of scores. The three means are $\bar{Y}_{A_1} = 15$, $\bar{Y}_{A_2} = 6$, and $\bar{Y}_{A_3} = 9$. The grand mean of all three conditions, obtained by taking the average of the three treatment means, is $\bar{Y}_T = (15 + 6 + 9)/3 = 10$. (I will illustrate the calculations in the next section.) As explained previously, we cannot conclude that any differences among the group means represent the "real" effects of the different experimental treatments: The differences may have resulted from experimental error, the short-term siding of uncontrolled sources of variability with one treatment condition or another. You saw that the solution to this problem is to compare the differences among the group means against the differences obtained from subjects within each of the individual groups. Let us see how this is accomplished.

For the moment, I will focus on the worst score in condition a_2, which is two correct responses on the objective test given following the presentation of the essay. I will represent this score as $Y_{2,5}$, where the first subscript specifies the level of factor A (a_2 in this case) and the second subscript indicates the subject's ordinal position in the original listing of the scores. This subject happens to be the fifth score in level a_2 as listed in Table 2–1. (I will discuss the notational system in the next section.)

Consider now the deviation of this score ($Y_{2,5}$) from the grand mean $\bar{Y}_T$. This deviation ($Y_{2,5} - \bar{Y}_T$) is represented geometrically at the bottom of Fig. 2–1 as the distance between the two vertical lines drawn from this score ($Y_{2,5} = 2$) on the left and the grand mean ($\bar{Y}_T = 10$) on the right, respectively. Consider next the vertical line drawn through the group mean for condition a_2 ($\bar{Y}_{A_2} = 6$). From the figure, it is obvious that this deviation is made up of *two components*. One component consists of the deviation of the score from the mean of the group from which it was selected, that is, $Y_{2,5} - \bar{Y}_{A_2}$, the component deviation on the left. The other component consists of the deviation of the group mean from the grand mean, that is, $\bar{Y}_{A_2} - \bar{Y}_T$, the component on the right. This relationship may be written as

$$Y_{2,5} - \bar{Y}_T = (\bar{Y}_{A_2} - \bar{Y}_T) + (Y_{2,5} - \bar{Y}_{A_2})$$

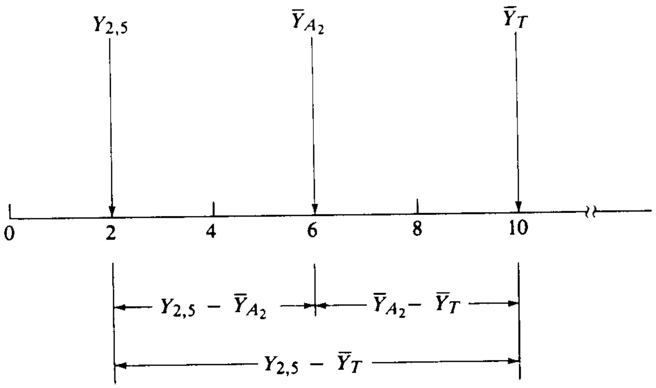

FIGURE 2–1 Geometric representation of the component deviations.

We can give each of the three deviations a name:

$Y_{2,5} - \overline{Y}_T$ is called the **total deviation.**

$\overline{Y}_{A_2} - \overline{Y}_T$ is called the **between-groups deviation.**

$Y_{2,5} - \overline{Y}_{A_2}$ is called the **within-groups deviation.**

This subdividing of the total deviation—or **partitioning,** as it is called—is illustrated with actual numbers in Table 2–2 under the heading, Level a_2. To illustrate again for the fifth subject, $Y_{2,5}$, we have

$$Y_{2,5} - \overline{Y}_T = (\overline{Y}_{A_2} - \overline{Y}_T) + (Y_{2,5} - \overline{Y}_{A_2})$$

Substituting for symbols ($Y_{2,5} = 2$, $\overline{Y}_{A_2} = 6$, and $\overline{Y}_T = 10$), we find

$$2 - 10 = (6 - 10) + (2 - 6)$$
$$-8 = (-4) + (-4)$$

Table 2–2 summarizes these calculations for each of the five subjects in this group. A similar partitioning can be conducted for each of the subjects in the other two groups. These are also summarized in the table.

Thus, you have seen that the score for each subject in an experiment can be expressed as a deviation from the grand mean and that this deviation can be partitioned into two components, a between-groups deviation and a within-groups deviation. These two component deviations are what we have been after, a quantity that will reflect treatment effects in the population (the between-groups deviation), in addition to experimental error, and a quantity that will reflect experimental error alone (the within-groups deviation). You will next see how these deviations are translated into measures of variability.

Table 2–2 Analysis of Component Deviation Scores

SCORE	DEVIATIONS				
Y	Total $(Y - \overline{Y}_T)$	= =	Between $(\overline{Y}_{A_i} - \overline{Y}_T)$	+ +	Within $(Y - \overline{Y}_{A_i})$
			Level a_1		
16	(6)	=	(5)	+	(1)
18	(8)	=	(5)	+	(3)
10	(0)	=	(5)	+	(−5)
12	(2)	=	(5)	+	(−3)
19	(9)	=	(5)	+	(4)
			Level a_2		
4	(−6)	=	(−4)	+	(−2)
6	(−4)	=	(−4)	+	(0)
8	(−2)	=	(−4)	+	(2)
10	(0)	=	(−4)	+	(4)
2	(−8)	=	(−4)	+	(−4)
			Level a_3		
2	(−8)	=	(−1)	+	(−7)
10	(0)	=	(−1)	+	(1)
9	(−1)	=	(−1)	+	(0)
13	(3)	=	(−1)	+	(4)
11	(1)	=	(−1)	+	(2)

2.3 SUMS OF SQUARES: DEFINING OR CONCEPTUAL FORMULAS

To evaluate the null hypothesis, it is necessary to transform the between-groups and within-groups deviations into more useful quantities, namely, **variances.** For this reason the statistical analysis involving the comparison of variances—in this case between-groups and within-groups variances—reflecting different sources of variability is called the **analysis of variance.** A variance is defined as follows:

$$\text{variance} = \frac{\text{sum of the squared deviations from the mean}}{\text{degrees of freedom}}$$

The quantity specified in the numerator, the **sum of the squared deviations from the mean,** usually shortened to **sum of squares** and abbreviated ***SS***, reflects the degree to which the numbers in a set vary among themselves. When there is no variability, that is, when the numbers are all the same, each deviation from the mean will be zero (because the mean is equal to each number in the set), and the sum of the squared deviations (as well as the variance) will equal zero. On the other hand, when the numbers in a set are *different*, they spread out from the mean and the sum of the squared deviations becomes a positive value. As the spread increases, the deviations from the mean increase and the *SS* increases as well.

The quantity in the denominator, **degrees of freedom (*df*),** is approximately equal to the number of numbers in the set. This means, therefore, that the variance is basically an *average* of the squared deviations. I will consider variances and the concept of degrees of freedom in the next chapter. Our concern now is the calculation of component sums of squares.

Let's return to the component deviations developed in Sec. 2.2. You will recall that the deviation of each subject from the grand mean of the experiment can be divided into a between-group portion and a within-group portion. A similar additive relationship holds if we square the three deviations for each subject and then sum the squares over all of the subjects in the experiment. This important relationship may be stated as

$$SS_{total} = SS_{between\ groups} + SS_{within\ groups} \qquad (2\text{–}1)$$

Translated into the numerical example from Table 2–1, Eq. (2–1) reads, "The sum of the squared deviations of all 15 subjects from $\bar{Y}_T$ may be broken down into two components, one obtained by summing all the squared deviations between individual group means and $\bar{Y}_T$ and the other by summing all the squared deviations of subjects from their respective group means." You will now see how these words are translated into formulas and numerical values.

Total Sum of Squares

The basic ingredients in the total sum of squares SS_T are the total deviations—that is, the deviations of all the scores in the experiment from the grand mean, $\bar{Y}_T$. The SS_T is formed by squaring the total deviation for each subject and summing the squares of the total deviations. In symbols,

$$SS_T = \Sigma (Y - \bar{Y}_T)^2 \qquad (2\text{–}2)$$

As you undoubtedly know, the capital Greek letter Σ, sigma, is read, "the sum of . . . ." Thus, $\Sigma (Y - \bar{Y}_T)^2$ is read, "the sum of the deviations formed by subtracting the grand mean of the Y scores from all the Y scores in the experiment." We can calculate SS_T for the numerical example we have been considering by squaring each of the 15 total deviations presented in Table 2–2. More specifically,

$$SS_T = \Sigma (Y - \bar{Y}_T)^2$$
$$= (6)^2 + (8)^2 + \ldots + (3)^2 + (1)^2 = 380$$

Between-Groups Sum of Squares

You saw in Fig. 2–1 that one of the components of a subject's total deviation is the deviation of the subject's group mean from the grand mean $(\bar{Y}_A - \bar{Y}_T)$. If we square this component and sum the squares for all the subjects in the experiment, we will obtain the between-groups SS. (I will refer to this quantity as the SS_A, indicating that this sum of squares is based on the deviations involving the A means.) From the between-group deviations listed in Table 2–2,

$$SS_A = (5)^2 + (5)^2 + \ldots + (-1)^2 + (-1)^2 = 210$$

Alternatively, we can express this formula differently by taking advantage of the

fact that the between-group deviation is the same for all subjects in any given group. In this example, the deviation is 5 for a_1, −4 for a_2, and −1 for a_3. We can obtain the same quantity by squaring these three between-groups deviations, multiplying the squared deviations by the number of subjects in each group (n), and then summing the three products. That is,

$$SS_A = n(\bar{Y}_{A_1} - \bar{Y}_T)^2 + n(\bar{Y}_{A_2} - \bar{Y}_T)^2 + n(\bar{Y}_{A_3} - \bar{Y}_T)^2$$

In general, the equation becomes

$$SS_A = \Sigma n(\bar{Y}_A - \bar{Y}_T)^2 \qquad (2\text{–}3)$$

which may be simplified by placing n, the sample size, to the left of the summation sign; that is, it is simpler to square and sum the deviations first and then to multiply by n. With this example,

$$SS_A = n\Sigma (\bar{Y}_A - \bar{Y}_T)^2$$
$$= 5[(5)^2 + (-4)^2 + (-1)^2]$$
$$= 5(25 + 16 + 1) = 5(42) = 210$$

Within-Groups Sum of Squares

The final sum of squares is the within-groups sum of squares, denoted by $SS_{S/A}$. The subscript S/A is read "S within A" and stresses the fact that we are dealing with the deviation of subjects from their own group means. As illustrated in Fig. 2–1, the basic deviation involved in the definition of the $SS_{S/A}$ is expressed by $Y - \bar{Y}_A$. As a first step, we can obtain a sum of squares for *each group* by using these within-group deviations. From Table 2–2,

$$SS_{S/A_1} = (1)^2 + (3)^2 + (-5)^2 + (-3)^2 + (4)^2 = 60$$
$$SS_{S/A_2} = (-2)^2 + (0)^2 + (2)^2 + (4)^2 + (-4)^2 = 40$$
$$SS_{S/A_3} = (-7)^2 + (1)^2 + (0)^2 + (4)^2 + (2)^2 = 70$$

In the analysis of variance, we will average the different within-group variances to obtain a more stable estimate of experimental error. As a first step, then, we will want to add together the separate within-group sums of squares, a process often referred to as **pooling**. In this example,

$$SS_{S/A} = SS_{S/A_1} + SS_{S/A_2} + SS_{S/A_3}$$
$$= 60 + 40 + 70 = 170$$

The equation for the pooled sum of squares may be expressed as follows:

$$SS_{S/A} = \Sigma (Y - \bar{Y}_A)^2 \qquad (2\text{–}4)$$

where it is understood that the summation sign refers to all of the within-group deviations.

2.4 SUMS OF SQUARES: COMPUTATIONAL FORMULAS

Although the defining formulas for the three sums of squares preserve the logic by which the component deviations are derived, we usually calculate the sums of squares with formulas that are equivalent algebraically but much simpler computationally.

Notation

Before presenting the different formulas, I should say a few words about the notational system adopted for this book. The basic job of any notational system is to express unambiguously the arithmetic operations in the most complex of designs as well as in the simplest. The system used in this book is designed specifically for the analysis of variance and to facilitate the calculation of sums of squares. You probably will not fully appreciate the pedagogical advantages of this system until we turn to the analysis of more complicated designs.

There are three major advantages of this system. First, the system uses different capital letters or different combinations of letters to designate basic quantities needed to calculate sums of squares. The confusion generated by the more usual notational systems—which consist of subscripts, parentheses, and multiple summation signs with subscripts and superscripts—is avoided, and the differences in the arithmetical operations are emphasized and made more distinct. Second, the system is designed to facilitate the reiterative computational sequences that are a part of the various calculations. As you will see in Chap. 10, the notational system works in conjunction with a general analysis scheme that can be applied to nearly all the designs I will consider in this book. Finally, the system adopted for this edition is compatible with the notation introduced by Keppel and Zedeck (1989), in which analysis of variance and multiple regression are applied to the analysis of the same experimental designs.[1]

A notational system is essentially a code. The symbols constitute a shorthand for specifying the operations to be performed on the data of an experiment. In the analysis of the completely randomized single-factor design, we need to designate only *three* basic quantities: the individual scores or observations (the *raw data*), the sum of these scores for each treatment condition (the *treatment sums* or *subtotals*), and the sum of all the scores or observations (the *grand sum* or *grand total*).

The Individual Scores. Each subject provides a single numerical value that reflects his or her performance on the response measure. This basic score or observation is designated by a single capital letter, Y. Table 2–3 illustrates the use of this notation with numbers obtained from the numerical example of Table 2–1. I have used numerical subscripts so that each score is specified uniquely. The first number in the subscript designates the level of factor A and the second number designates the particular subject within that treatment condition. To refer to a score without specifying any particular one, I use Y without a subscript, or occasionally, with letters as subscripts, Y_{ij}. Technically, then, the subscript i refers to the levels of the independent variable, factor A, and can take on values of $i = 1, 2, \ldots, a$, and the subscript j can take on values of $j = 1, 2, \ldots, n$.

[1]For the benefit of those who are familiar with earlier editions of this book, the major changes are the use of Y to designate an individual score, of n to represent sample size, and of more standard symbols for means. I have retained the letter coding for sums and computational formulas.

Table 2–3 A Summary of the Notational System

	Level a_1	Level a_2	Level a_3	Total
	$Y_{1,1} = 16$	$Y_{2,1} = 4$	$Y_{3,1} = 2$	
	$Y_{1,2} = 18$	$Y_{2,2} = 6$	$Y_{3,2} = 10$	
	$Y_{1,3} = 10$	$Y_{2,3} = 8$	$Y_{3,3} = 9$	
	$Y_{1,4} = 12$	$Y_{2,4} = 10$	$Y_{3,4} = 13$	
	$Y_{1,5} = 19$	$Y_{2,5} = 2$	$Y_{3,5} = 11$	
Sums	$A_1 = 75$	$A_2 = 30$	$A_3 = 45$	$T = \Sigma A = 150$
No. of Observations	$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$(a)(n) = 15$
Means	$\bar{Y}_{A_1} = 15$	$\bar{Y}_{A_2} = 6$	$\bar{Y}_{A_3} = 9$	$\bar{Y}_T = 10$

The Treatment Sums. As a first step in the analysis, we will calculate the sums of the scores in each of the treatment groups. These subtotals, or treatment sums, are designated by a capital A to stand for the sums of the scores obtained under the levels of factor A. A numerical subscript permits the specification of the treatment sum for a particular treatment condition. The meaning of this symbol and the subscript are illustrated in Table 2–3. To designate a treatment sum without specifying any sum in particular, I will use A without a subscript, or occasionally with an i as a subscript: A_i. As a numerical example, the respective treatment sums for the three groups at levels a_1, a_2, and a_3 are

$$A_1 = 16 + 18 + 10 + 12 + 19 = 75$$
$$A_2 = 4 + 6 + 8 + 10 + 2 = 30$$
$$A_3 = 2 + 10 + 9 + 13 + 11 = 45$$

To obtain a numerical summary of the outcome of the experiment, we use the treatment sums to calculate the **treatment means**, $\bar{Y}_A$, by dividing each treatment sum by the number of scores in each of the conditions. In this book, I use a lower-case n to designate **sample size**, that is, the number of subjects in a treatment condition.[2] Thus, the formula for a treatment mean is

$$\bar{Y}_A = \frac{A}{n} \quad (2\text{–}5)$$

[2]In most of the designs I will consider, equal sample sizes (n) are used in the treatment conditions. Although this represents a special case, it subsumes most of the experiments conducted in the behavioral sciences. The analysis of unequal sample sizes is presented in Chap. 13.

To refer to specific treatment means, I use the number subscript. From the data in Table 2–3,

$$\bar{Y}_{A_1} = \frac{A_1}{n} = \frac{75}{5} = 15.00$$

$$\bar{Y}_{A_2} = \frac{A_2}{n} = \frac{30}{5} = 6.00$$

$$\bar{Y}_{A_3} = \frac{A_3}{n} = \frac{45}{5} = 9.00$$

The Grand Sum. I designate the **grand sum** of the scores—that is, the sum of all the scores in the experiment—as T. Computationally, T may be calculated by summing the entire set of Y scores or by summing the treatment subtotals (A). Expressing these operations in symbols,

$$T = \Sigma Y = \Sigma A$$

(The first summation, ΣY, is read, "the sum of all the Y scores," and ΣA is read, "the sum of all the A treatment subtotals.") When we translate this sum into a mean (the **grand mean**), the mean is designated $\bar{Y}_T$. We calculate $\bar{Y}_T$ as we would any arithmetic mean, by dividing the sum of the scores by the number of scores. This number can be calculated by multiplying the number of scores in each treatment group (n) by the number of treatment groups, which we will designate a. In symbols, the total number of scores is $a \times n$, and the grand mean is

$$\bar{Y}_T = \frac{T}{(a)(n)} \tag{2–6}$$

From the numbers in Table 2–3, where $T = 75 + 30 + 45 = 150$, $a = 3$ treatment groups, and $n = 5$ subjects per group,

$$\bar{Y}_T = \frac{150}{(3)(5)} = 10.00$$

Basic Ratios

Each of the sums of squares is calculated by adding and subtracting special quantities that I call **basic ratios**. Basic ratios represent a common step in the computational formulas for sums of squares in the analysis of variance. All basic ratios have the same form and are calculated in a series of simple arithmetic steps. For the present design, there are three basic ratios, one involving the individual observations (Y), another involving the treatment subtotals (A), and the other involving the grand total (T).

The *numerator term* for any basic ratio involves two arithmetic steps:

1. The initial *squaring* of a set of quantities—the Y's, the A's, or T
2. *Summing* the squared quantities if more than one is present

To calculate the three required numerators, we simply conduct the two operations of squaring and then summing on each of the Y's, A's, and T, that is, ΣY^2, ΣA^2, and T^2, respectively. To illustrate, the first quantity indicates that we are to sum the squared scores for all the subjects in the experiment; using the data from Table 2–3, we have

$$\Sigma Y^2 = (16)^2 + (18)^2 + \ldots + (13)^2 + (11)^2 = 1{,}880$$

The second quantity specifies the sum of all the squared treatment sums; using the subtotals from Table 2–3, we get

$$\Sigma A^2 = (75)^2 + (30)^2 + (45)^2 = 5{,}625 + 900 + 2{,}025 = 8{,}550$$

The final quantity involves the grand total, which merely needs to be squared since there is only one such quantity in an experiment. From Table 2–3,

$$T^2 = (150)^2 = 22{,}500$$

The *denominator term* for any basic ratio is found by applying a simple rule that depends on the particular term appearing in the numerator:

> **Whatever the term, we divide by the number of scores that contributed to that term.**

For example, if the term is T, we divide by $(a)(n)$ because this is the number of scores that are actually summed to produce T. If the term is A, we divide by n because this is the number of scores that are summed to produce any one of the A treatment sums. Finally, if the term is Y, we divide by one, because a Y score is based on a *single* observation; this is equivalent, of course, to not dividing at all—that is, $\Sigma Y^2/1 = \Sigma Y^2$.

We are now ready to consider the formulas for the three basic ratios. For convenience in presenting computational formulas, I designate each basic ratio by a "bracket term"—a notational symbol consisting of a pair of brackets enclosing the letter appearing in the numerator of the basic ratio, that is, [Y] for the Y scores, [A] for the treatment sums, and [T] for the grand total. The formulas for the three basic ratios are

$$[T] = \frac{T^2}{(a)(n)} \tag{2–7}$$

$$[A] = \frac{\Sigma A^2}{n} \tag{2–8}$$

$$[Y] = \Sigma Y^2 \tag{2–9}$$

Applying these new formulas to the partial answers we have already calculated, we find

$$[T] = \frac{22{,}500}{(3)(5)} = 1{,}500.00$$

$$[A] = \frac{8{,}550}{5} = 1{,}710.00$$

$$[Y] = 1{,}880$$

Sums of Squares

All that remains is to specify how the three basic ratios are combined to produce the three required sums of squares, SS_T, SS_A, and $SS_{S/A}$. This final step is accomplished simply by noting that each of the three deviations appearing in the defining formulas presented in Sec. 2.3 separately indicates how the basic ratios are combined to form the appropriate computational formulas. To illustrate, the total sum of squares is based on the following deviation:

$$Y - \bar{Y}_T$$

The computational formula combines the two basic ratios identified by this deviation, $[Y]$ and $[T]$, as follows:

$$SS_T = [Y] - [T] \qquad (2\text{–}10)$$

or more completely,

$$SS_T = \Sigma Y^2 - \frac{T^2}{(a)(n)} \qquad (2\text{–}11)$$

Substituting the quantities we calculated in the last section, we find

$$\begin{aligned} SS_T &= [Y] - [T] \\ &= 1{,}880 - 1{,}500.00 = 380.00 \end{aligned}$$

The treatment sum of squares, SS_A, is based on the deviation of the treatment means from the grand mean:

$$\bar{Y}_A - \bar{Y}_T$$

The computational formula combines the two ratios identified by this deviation as follows:

$$SS_A = [A] - [T] \qquad (2\text{–}12)$$

or more fully,

$$SS_A = \frac{\Sigma A^2}{n} - \frac{T^2}{(a)(n)} \qquad (2\text{–}13)$$

Substituting the quantities calculated previously, we find

$$\begin{aligned} SS_A &= [A] - [T] \\ &= 1{,}710.00 - 1{,}500.00 = 210.00 \end{aligned}$$

Finally, the within-groups sum of squares is based on the deviation of individual observations from the relevant treatment mean, that is,

$$Y - \bar{Y}_A$$

The computational formula combines the two ratios identified by these deviations as follows:

$$SS_{S/A} = [Y] - [A] \qquad (2\text{–}14)$$

or

$$SS_T = \Sigma Y^2 - \frac{\Sigma A^2}{n} \qquad (2\text{–}15)$$

Substituting the quantities from the last section, we obtain

$$\begin{aligned} SS_{S/A} &= [Y] - [A] \\ &= 1{,}880 - 1{,}710.00 = 170.00 \end{aligned}$$

As a computational check and as a demonstration of the relationship among these three sums of squares, we will apply Eq. (2–1) to these calculations:[3]

$$SS_T = SS_A + SS_{S/A} = 210.00 + 170.00 = 380.00$$

Comment

You can verify for yourself that the computational formulas produced exactly the same answers as those obtained with the defining formulas previously. The only time different answers are found is when rounding errors introduced in calculating the treatment means and the grand mean become magnified in squaring the deviations, but even then the differences will be small. Although we will focus almost entirely on the computational versions of the formulas because they are easy to use and can be generated by a simple set of rules, you should keep in mind that they are

[3] A complete check of all our calculations may be obtained in a number of ways. One obvious method is to perform the analysis again or, perhaps better still, to coax another person to go through the calculations independently. An alternative method is to add a constant, say, 1, to each Y score (that is, to use $Y + 1$) and to repeat the complete analysis. For example, the original scores in level a_1 (16, 18, 10, 12, and 19) would become 17, 19, 11, 13, and 20, respectively. If you have made no error in either set of calculations, you should end up with *identical* sums of squares in the two analyses. The addition of a constant does not change the basic *deviations*, on which the sums of squares are based, but it does change the actual numbers entering into the calculations when we use the computational formulas presented in this section.

equivalent to the defining versions, which reflect quite directly the logic behind the derivation of the sums of squares required for the analysis of variance.

2.5 SUMMARY

I considered first some of the logic underlying the process of hypothesis testing in a design where each subject serves in only one treatment condition. By way of summary, we can describe hypothesis testing as consisting of a contrast between two sets of differences. One of these sets is obtained from a comparison involving differences among the treatment means; these differences are often referred to as *between-groups* differences. The other set is obtained from a comparison involving differences among subjects receiving the same treatment within a treatment group; these differences are called *within-groups* differences. I argued that the between-groups differences are the result of the combined effects of the experimental treatment and of experimental error, whereas the within-groups differences represent the influence of experimental error alone. You saw that the comparison ratio,

$$\frac{\text{between-groups differences}}{\text{within-groups differences}}$$

provides a numerical index that is "sensitive" to the presence of treatment effects in the population. That is, with no treatment effects, the long-run expectation is that the ratio will approximate 1.0 since the treatment effects will be zero and we will be dividing one estimate of experimental error by the other. On the other hand, whenever there are treatment effects, the expectation is that the ratio will be greater than 1.0.

The statistical hypothesis we test, the null hypothesis, specifies the *absence* of treatment effects in the population. With the help of statistical tables and a set of decision rules, neither of which I have described yet, we can decide whether or not it is reasonable to reject the null hypothesis. If we reject the null hypothesis, we accept the alternative statistical hypothesis, which specifies the presence of treatment effects in the population. If we fail to reject the null hypothesis, essentially we conclude that the independent variable produced no systematic differences in the experiment.

The remainder of the chapter focused on the translation of these ideas into actual measures of variability based on the data of a single-factor experiment. You saw that between-groups and within-groups differences can be expressed as deviations from different means. I began by considering the deviation of Y scores from $\overline{Y}_T$ and established the fact that for each observation this deviation may be divided into a between-group deviation $\overline{Y}_A - \overline{Y}_T$ and a within-groups deviation $Y - \overline{Y}_A$. You saw that the defining formulas for the corresponding sums of squares were developed directly from these deviations. Our actual calculations, however, are performed with computational formulas, which are considerably easier to use and can be formed by using a set of simple rules. In the next chapter, I will complete the steps in the statistical evaluation of the null hypothesis.

2.6 EXERCISES[4]

1. In an experiment involving $a = 5$ treatment conditions, the following measures were obtained:

a_1	a_2	a_3	a_4	a_5
13	7	12	10	13
9	4	11	12	6
8	4	4	9	14
7	1	9	7	12
8	10	5	15	13
6	7	10	14	10
6	5	2	10	8
7	9	8	17	4
6	5	3	14	9
10	8	6	12	11

a. Calculate the sum of the scores and the sum of the squared scores for each of the treatment groups.
b. Calculate the treatment means.
c. Calculate the sums of squares for each of the sources of variability normally identified in the analysis of variance, that is, SS_A, $SS_{S/A}$, and SS_T. Reserve this information for problem 3 in the exercises for Chap. 3.

[4]The answers to this problem are found in Appendix B.