As knowledge and clinical theory have developed, clinical researchers have proposed more complex research questions, necessitating the use of elaborate multilevel and multifactor experimental designs. The **analysis of variance (ANOVA)** is a powerful analytic tool for analyzing such designs, where three or more conditions or groups are compared. The analysis of variance is used to determine if the observed differences among a set of means are greater than would be expected by chance alone. The ANOVA is based on the *F* statistic, which is similar to *t* in that it is a ratio of between-groups treatment effects to within-group variability. The test can be applied to independent groups or repeated measures designs.*

The purpose of this chapter is to describe the application of the analysis of variance for a variety of experimental research designs. An introduction to the basic concepts underlying analysis of variance is most easily addressed in the context of a single-factor experiment (one independent variable) with independent groups. We then follow with discussions of more complex models, including factorial designs and repeated measures designs.

ANALYSIS OF VARIANCE FOR INDEPENDENT SAMPLES: ONE-WAY CLASSIFICATION

In a single-factor experiment, the one-way analysis of variance is applied when three or more independent group means are compared. The descriptor "one-way" indicates that the design involves one independent variable, or factor, with three or more levels.

---

*As with all parametric tests, the ANOVA is based on the assumption that samples are drawn randomly from normally distributed populations with equal variances. Tests for homogeneity of variance can be performed to validate the latter assumption. With samples of equal size, the analysis of variance is considered "robust" in that reasonable departures from the assumptions of normality and homogeneity will not seriously affect the validity of inferences drawn from the data.[1] With unequal sample sizes, gross violations of homogeneity of variance can increase the chance of Type I error. In such cases, a nonparametric analysis of variance can be applied (see Chapter 22), or data can be transformed to a different scale that improves homogeneity of variance within the sample distribution (see Appendix D).

Although the ANOVA can be applied to two-group comparisons, the t-test is generally considered more efficient for that purpose.†

Statistical Hypotheses

The null hypothesis for a one-way multilevel study states that there is no significant difference among the group means, represented by

$$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$$

where k is the number of groups or levels of the independent variable. The alternative hypothesis (H_1) states that *at least two means* will differ.

Sums of Squares

In the last chapter we established that mean differences can be evaluated using a statistical ratio that relates the treatment effect to experimental error. The analysis of variance uses the same process, except that the ratio must now account for the relationships among several means. The F-test (named for Sir Ronald Fisher, who developed the test) is used to determine how much of the total observed variability in scores can be explained by differences among several treatment means and how much is attributable to unexplained differences among subjects. To analyze this variability with several groups, we must refer to the concept of **sum of squares (SS)**, introduced in Chapter 17. The sum of squares is calculated by subtracting the sample mean from each score ($X - \overline{X}$), squaring those values, and taking their sum ($SS = \Sigma(X - \overline{X})^2$). The larger the sum of squares, the greater the variability of scores within a sample.

Example

To illustrate how this concept is applied to analysis of variance, consider a hypothetical study of the effect of using different modalities for 10 days to gain pain-free range of motion (ROM) in patients with tendonitis. Through random assignment, we create four independent groups: one to get ultrasound (US), a second to get ice, a third to get massage, and a fourth group to serve as a control (see Figure 20.1). We use a lowercase n to indicate the number of subjects in each group ($n = 11$) and an uppercase N to represent the total number of subjects in the study ($N = 44$). The independent variable, type of modality, has four levels ($k = 4$). Therefore, this is a single-factor, multilevel design. The dependent variable is elbow ROM, measured in degrees. Hypothetical data for this study are reported in Table 20.1A.

---

†The results of a t-test and analysis of variance with two groups will be the same. The t-test is actually a special case of the analysis of variance, with the relationship $F = t^2$.

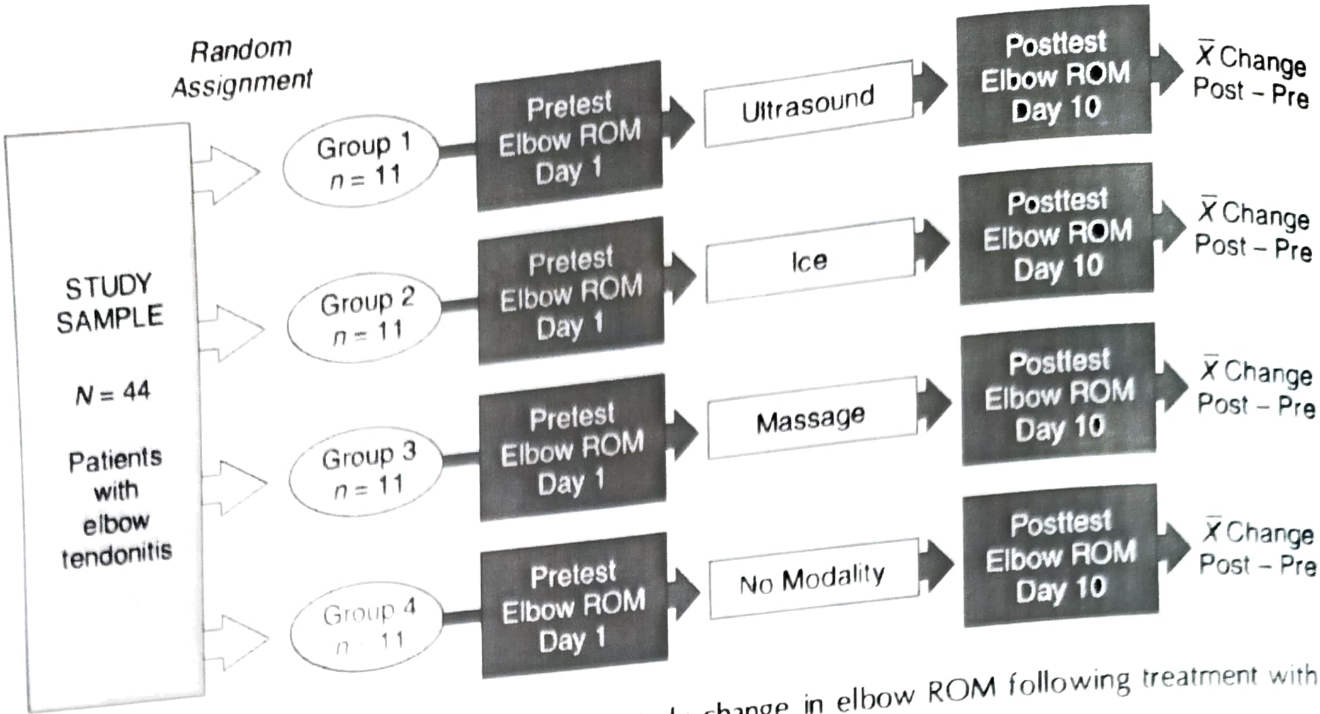

FIGURE 20.1 One-way multi-group design to study change in elbow ROM following treatment with different modalities in patients with tendonitis.

Total Sum of Squares

To estimate the total variability in these data, consider the set of 44 scores as one total sample, ignoring group assignment. We can calculate a mean for this total sample, called the **grand mean,** $\overline{X}_G$, around which all 44 scores will vary. For the data in Table 20.1, the sum of all 44 scores is 1,638, and $\overline{X}_G = 37.23$. The sum of squares for this total sample ($\Sigma(X - \overline{X}_G)^2$) represents the deviations of each individual score from the grand mean. This *total sum of squares* (SS_t) reflects the *total variability* that exists within this set of 44 scores. This variability is illustrated in Figure 20.2A, showing the entire distribution of scores above and below the grand mean.

Partitioning Sum of Squares

As we have described before, total variability in a set of data can be attributed to two sources: a treatment effect (*between* the groups), and unexplained sources of variance, or **error variance**, among the subjects (*within* the groups). As its name implies, the analysis of variance partitions the total variance within a set of data (SS_t) into these two components. The *between-groups sum of squares* (SS_b) reflects the spread of group means around the grand mean. The larger this effect, the greater the separation between the groups. The within-groups or *error sum of squares* (SS_e) reflects the spread of scores within each group around the group mean, or the differences among subjects. In Figure 20.2B, we can see that the means for groups 1 and 2 are close together, and both appear separated from groups 3 and 4. The spread of scores in group 4 appears to be less than in the other groups.

Because hand calculations are complex, computer programs will most often be used to obtain results for an ANOVA. For those who like to see the math, computational

TABLE 20.1 ONE-WAY ANALYSIS OF VARIANCE FOR INDEPENDENT SAMPLES: CHANGE IN ELBOW ROM (IN DEGREES) FOLLOWING TREATMENT FOR TENDONITIS ($k = 4, N = 44$)

A. DATA

	Grp	ROM		Grp	ROM
1	1	23	23	3	47
2	1	54	24	3	49
3	1	52	25	3	29
4	1	33	26	3	33
5	1	48	27	3	45
6	1	52	28	3	29
7	1	58	29	3	43
8	1	31	30	3	19
9	1	43	31	3	34
10	1	47	32	3	27
11	1	45	33	3	33
12	2	44	34	4	19
13	2	52	35	4	14
14	2	53	36	4	23
15	2	52	37	4	14
16	2	33	38	4	36
17	2	46	39	4	29
18	2	56	40	4	37
19	2	42	41	4	22
20	2	43	42	4	19
21	2	29	43	4	18
22	2	48	44	4	35

	Group 1 US	Group 2 Ice	Group 3 Massage	Group 4 Control	Total
$\sum X$	486.00	498.00	388.00	266.00	1,638.00
n	11	11	11	11	44
$\bar{X}$	44.18	45.27	35.27	24.18	37.23

B. OUTPUT

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	❶ Sig.
LENGTH	.321	3	40	.810

ANOVA

LENGTH	Sum of Squares	df	Mean Square	F	❷ Sig.
Between Groups	3158.09	3	1052.70	11.89	.000
❸ Within Groups	3541.64	40	88.54		
Total	6699.73	43			

❶ As with the *t*-test, the Levene statistic indicates that there is no significant difference ($p = .810$) between the variances across the four groups.

❷ The probabilities associated with the *F* test do not distinguish between one and two-tailed tests. Because the probability is less than .05, we reject H_0 and conclude that there is a significant difference among groups.

❸ In different programs, the source of variance "Within Groups" may also be called "Error" or "Residual" variance.

formulae for calculating total, between-groups and error sums of squares are shown in Table 20.2.

The *F* Statistic

Degrees of Freedom

The total degrees of freedom (df_t) within a set of data will always be one less than the total number of observations, in this case $N - 1$. In our example, $N = 44$ and $df_t = 43$ The number of degrees of freedom associated with the between-groups variability (df_b