Chapter 8
Analysis of Variance

Learning Objectives

8.1 Illustrate how analysis of variance can be used to make comparisons among three or more sample means

8.2 Calculate and interpret the total, within-in groups, and between-groups sums of squares

8.3 Illustrate how the mean square overcomes the problem of sample size affecting the calculation of variation

8.4 Analyze reasons as to why the larger the *F* ratio, the more likely we are to reject the null hypothesis

8.5 Calculate Tukey's HSD as a useful test for investigating the multiple comparison of means

8.6 Examine the influence of two factors or independent variables using two-way analysis of variance

8.7 Identify the requirements that need to be considered before the analysis of variance is made

Introduction

Blacks versus whites, males versus females, and liberals versus conservatives represent the kind of two-sample comparisons that occupied our attention earlier in the course. Yet social reality cannot always be conveniently sliced into two groups; respondents do not always divide themselves in so simple a fashion.

As a result, the social researcher often seeks to make comparisons among three, four, five, or more samples or groups. To illustrate, he or she may study the influence of:

- Racial/ethnic identity (black, Latino, white, or Asian) on job discrimination
- Degree of economic deprivation (severe, moderate, or mild) on juvenile delinquency, or
- Subjective social class (upper, middle, working, or lower) on achievement motivation

You may wonder whether we can use a *series* of *t* ratios to make comparisons among three or more sample means. Suppose, for example, we want to test the influence of subjective social class on achievement motivation. Can we simply compare all possible pairs of social class in terms of our respondents' levels of achievement motivation and obtain a *t* ratio for each comparison? Using this method, four samples of respondents would generate six paired combinations for which six *t* ratios must be calculated:

1. Upper class versus middle class
2. Upper class versus working class
3. Upper class versus lower class
4. Middle class versus working class
5. Middle class versus lower class
6. Working class versus lower class

Not only would the procedure of calculating a series of *t* ratios involve a good deal of work, but it has a major statistical limitation as well. This is because it increases the probability of making a Type I error—the error of rejecting the null hypothesis when it is true and should be retained.

Recall that the social researcher is generally willing to accept a 5% risk of making a Type I error (the .05 level of significance). He or she therefore expects that *by chance alone* 5 out of every 100 sample mean differences will be large enough to be considered as significant. The more statistical tests we conduct, however, the more likely we are to get statistically significant findings by sampling error (rather than by a true population difference) and hence to commit a Type I error. When we run a large number of such tests, the interpretation of our result becomes problematic. To take an extreme example: How would we interpret a significant t ratio out of 1,000 such comparisons made in a particular study? We know that at least a few large mean differences can be expected to occur simply on the basis of sampling error. Let's look at a more typical example.

Show Example

Suppose a researcher wished to survey and compare voters in eight regions of the country (New England, Middle Atlantic, South Atlantic, Midwest, South, Southwest, Mountain, and Pacific) on their opinions about the President. Comparing the regional samples would require 28 separate t ratios (New England versus Middle Atlantic, New England versus South Atlantic, Middle Atlantic versus South Atlantic, and so on). Out of 28 separate tests of difference between sample means, each with a .05 level of significance, 5% of the 28 tests—between one and two (1.4 to be exact)—would be expected to be significant due to chance or sampling error alone.

- Suppose that from the 28 different t ratios, the researcher obtains two t ratios (New England versus South, and Middle Atlantic versus Mountain) that are significant.
- How should the researcher interpret these two significant differences?
- Should he go out on a limb and treat both as indicative of real population differences?
- Should he play it safe by maintaining that both could be the result of sampling error and go back to collect more data?
- Should he, based on the expectation that one t ratio will come out significant by chance, decide that only one of the two significant t ratios is valid?
- If so, in which of the two significant t ratios should he have faith?
- The larger one?
- The one that seems more plausible?

Unfortunately, none of these solutions is particularly sound. The problem is that as the number of separate tests mounts, the likelihood of rejecting a true null hypothesis (Type I error) grows accordingly. Thus, while for each t ratio the probability of a Type I error may be .05, overall the probability of rejecting *any* true null hypothesis is far greater than .05.

To overcome this problem and clarify the interpretation of our results, we need a statistical test that holds Type I error at a constant level (for example, .05) by making a *single* overall decision as to whether a significant difference is present among the three, four, eight, or however many sample means we seek to compare. Such a test is known as the **analysis of variance**.

8.1: The Logic of Analysis of Variance

Objective: Illustrate how analysis of variance can be used to make comparisons among three or more sample means

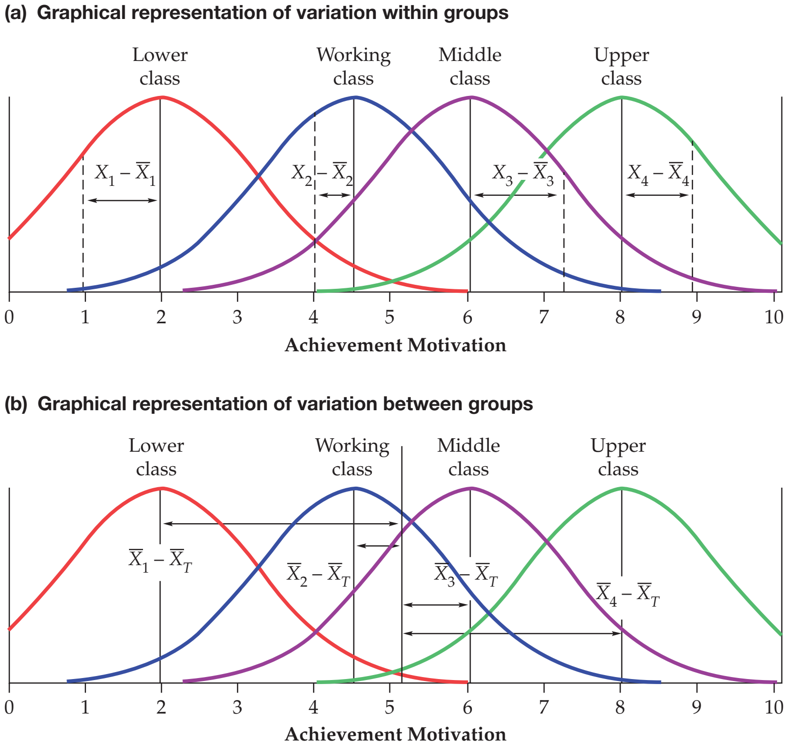
To conduct an analysis of variance (also called ANOVA, for short), we treat the total *variation* in a set of scores as being divisible into two components:

- The distance or deviation of raw scores from their group mean, known as **variation within groups**.
- The distance or deviation of group means from one another, referred to as **variation between groups**.

To examine variation within groups, the achievement–motivation scores of members of four social classes—(1) lower, (2) working, (3) middle, and (4) upper—are graphically represented in Figure 8.1(a), where X_1, X_2, X_3, and X_4 are any raw scores in their respective groups, and $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$, and $\bar{X}_4$ are the group means. In symbolic terms, we see that variation within groups refers to the deviations $X_1 - \bar{X}_1$, $X_2 - \bar{X}_2$, $X_3 - \bar{X}_3$, and $X_4 - \bar{X}_4$.

We can also visualize variation between groups. With the aid of Figure 8.1(b), we see that degree of achievement

Figure 8.1 Graphical Representation of Variation Within Groups and Between Groups



motivation varies by social class: The upper-class group has greater achievement motivation than the middle-class group which, in turn, has greater achievement motivation than the working-class group which, in its turn, has greater achievement motivation than the lower-class group. More specifically, we determine the overall mean for the total sample of all groups combined, denoted here as $\overline{X}_T$, and then compare each of the four group means to the total mean. In symbolic terms, we see that variation between groups focuses on the deviations $\overline{X}_1 - \overline{X}_T$, $\overline{X}_2 - \overline{X}_T$, $\overline{X}_3 - \overline{X}_T$, and $\overline{X}_4 - \overline{X}_T$.

The distinction between variation *within* groups and variation *between* groups is not peculiar to the analysis of variance. Although not named as such, we encountered a similar distinction in the form of the t ratio, wherein a difference *between* $\overline{X}_1$ and $\overline{X}_2$ was compared against the standard error of the difference $s_{\overline{X}_1 - \overline{X}_2}$, a combined estimate of differences *within* each group. That is,

$$t = \frac{\overline{X}_1 - \overline{X}_2 \quad \leftarrow \text{variation between groups}}{s_{\overline{X}_1 - \overline{X}_2} \quad \leftarrow \text{variation within groups}}$$

In a similar way, the analysis of variance yields an *F ratio*, whose numerator represents variation between the groups being compared, and whose denominator represents variation within these groups. As we shall see, the ***F* ratio** indicates the size of the variation between groups *relative* to the size of the variation within each group. As was true of the t ratio, the larger the F ratio (the larger the variation between groups relative to the variation within groups), the greater the probability of rejecting the null hypothesis and accepting the research hypothesis.

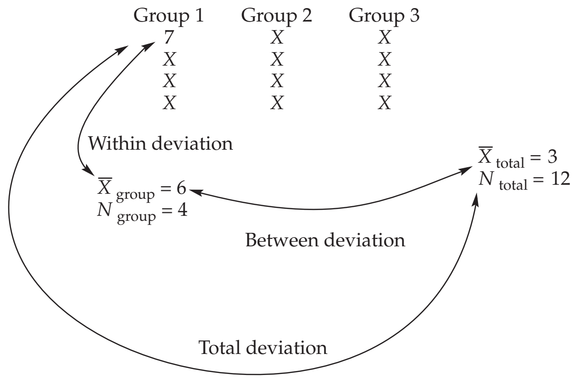
8.2: The Sum of Squares

Objective: Calculate and interpret the total, within-in groups, and between-groups sums of squares

At the heart of the analysis of variance is the concept of **sum of squares**, which represents the initial step for measuring total variation, as well as variation between and within groups. It may come as a pleasant surprise to learn that only the label "sum of squares" is new to us. The concept itself was introduced earlier in the course as an important step in the procedure for obtaining the variance. In that context, we learned to find the sum of squares by squaring the deviations from the mean of a distribution and adding these squared deviations together $\sum(X - \overline{X})^2$. This procedure eliminated minus signs while still providing a sound mathematical basis for the variance and standard deviation.

When applied to a situation in which groups are being compared, there is more than one type of sum of squares, although each type represents the sum of squared deviations from a mean. Corresponding to the distinction between total variation and its two components, we have the *total* sum of squares (SS_{total}), *between-groups* sum of squares ($SS_{between}$), and *within-groups* sum of squares (SS_{within}).

Figure 8.2 Analysis of Variance



Consider the hypothetical results shown in Figure 8.2. Note that only part of the data is shown to help us focus on the concepts of total, within-groups, and between-groups sums of squares.

The respondent with a 7 scored substantially higher than the total mean ($\overline{X}_{total} = 3$). His deviation from the total mean is $(X - \overline{X}_{total}) = 4$. Part of this elevated score represents, however, the fact that his group scored higher on average ($\overline{X}_{group} = 6$) than the overall or total mean ($\overline{X}_{total} = 3$). That is, the deviation of this respondent's group mean from the total mean is $(\overline{X}_{group} - \overline{X}_{total}) = 3$. After accounting for the group difference, this respondent's score remains higher than his own group mean. Within the group, his deviation from the group mean is $(X - \overline{X}_{group}) = 1$.

As we shall see very shortly, we can take these deviations (of scores from the total mean, between group means and the total mean, and of scores from their group means), square them, and then sum them to obtain SS_{total}, SS_{within}, and $SS_{between}$.

8.2.1: A Research Illustration

Let's consider a research situation in which each type of sum of squares might be calculated. Suppose a researcher is interested in comparing the degree of life satisfaction among adults with different marital statuses. She wants to know if single or married people are more satisfied with life and whether separated and divorced adults tend to have a more negative view of life. She selects at random five middle-aged adults from each of the following four categories: widowed, divorced, never married, and married. The researcher then administers to each of the 20 respondents a 40-item checklist designed to measure satisfaction with various aspects of life. The scale ranges from 0 for dissatisfaction with all aspects of life to 40 for satisfaction with all aspects of life.

Table 8.1 Satisfaction with Life by Marital Status

Widowed			Divorced		
X_1	$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	X_2	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
5	1	1	16	7	49
6	2	4	5	–4	16
4	0	0	9	0	0
5	1	1	10	1	1
0	–4	16	5	–4	16
$\Sigma X_1 = 20$	$\bar{X}_1 = \frac{20}{5} = 4$	$\Sigma(X_1 - \bar{X}_1)^2 = 22$	$\Sigma X_2 = 45$	$\bar{X}_2 = \frac{45}{5} = 9$	$\Sigma(X_2 - \bar{X}_2)^2 = 82$
Never Married			**Married**		
X_3	$X_3 - \bar{X}_3$	$(X_3 - \bar{X}_3)^2$	X_4	$X_4 - \bar{X}_4$	$(X_4 - \bar{X}_4)^2$
23	–1	1	19	–6	36
30	6	36	35	10	100
20	–4	16	15	–10	100
20	–4	16	26	1	1
27	3	9	30	5	25
$\Sigma X_3 = 120$	$\bar{X}_3 = \frac{120}{5} = 24$	$\Sigma(X_3 - \bar{X}_3)^2 = 78$	$\Sigma X_4 = 125$	$\bar{X}_4 = \frac{125}{5} = 25$	$\Sigma(X_4 - \bar{X}_4)^2 = 262$
		$\bar{X}_{\text{total}} = 15.5$			

The researcher sets up her hypotheses as follows:

Null hypothesis: $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$	*Marital status has no effect on degree of satisfaction with life.*
Research hypothesis: (some $\mu_i \neq \mu_j$)	*Marital status has an effect on degree of satisfaction with life.*

CALCULATING THE TOTAL SUM OF SQUARES The total sum of squares is defined as the sum of the squared deviation of every raw score from the total mean. By formula,

$$SS_{\text{total}} = \Sigma(X - \bar{X}_{\text{total}})^2$$

where X = any raw score

$\bar{X}_{\text{total}}$ = total mean for all groups combined

Using this formula, we subtract the total mean ($\bar{X}_{\text{total}}$) from each raw score (X), square the deviations that result, and then sum. Applying this formula to the data in Table 8.1, we obtain the following result:

$$\begin{aligned} SS_{\text{total}} &= (5 - 15.5)^2 + (6 - 15.5)^2 + (4 - 15.5)^2 \\ &\quad + (5 - 15.5)^2 + (0 - 15.5)^2 + (16 - 15.5)^2 \\ &\quad + (5 - 15.5)^2 + (9 - 15.5)^2 + (10 - 15.5)^2 \\ &\quad + (5 - 15.5)^2 + (23 - 15.5)^2 + (30 - 15.5)^2 \\ &\quad + (20 - 15.5)^2 + (20 - 15.5)^2 + (27 - 15.5)^2 \\ &\quad + (19 - 15.5)^2 + (35 - 15.5)^2 + (15 - 15.5)^2 \\ &\quad + (26 - 15.5)^2 + (30 - 15.5)^2 \\ &= (-10.5)^2 + (-9.5)^2 + (-11.5)^2 + (-10.5)^2 \\ &\quad + (-15.5)^2 + (.5)^2 + (-10.5)^2 + (-6.5)^2 \\ &\quad + (-5.5)^2 + (-10.5)^2 + (7.5)^2 + (14.5)^2 \\ &\quad + (4.5)^2 + (4.5)^2 + (11.5)^2 + (3.5)^2 \\ &\quad + (19.5)^2 + (.5)^2 + (10.5)^2 + (14.5)^2 \\ &= 110.25 + 90.25 + 132.25 + 110.25 + 240.25 \\ &\quad + .25 + 110.25 + 42.25 + 30.25 + 110.25 \\ &\quad + 56.25 + 210.25 + 20.25 + 20.25 + 132.25 \\ &\quad + 12.25 + 380.25 + .25 + 110.25 + 210.25 \\ &= 2{,}129 \end{aligned}$$

CALCULATING THE WITHIN-GROUPS SUM OF SQUARES The within-groups sum of squares is the sum of the squared deviations of every raw score from its group mean. By formula,

$$SS_{\text{within}} = \Sigma(X - \bar{X}_{\text{group}})^2$$

where X = any raw score

$\bar{X}_{\text{group}}$ = mean of the group containing the raw score

Using this formula, we subtract the group mean ($\bar{X}_{\text{group}}$) from each raw score (X), square the deviations that result, and then sum. Applying this formula to the data in Table 8.1, we obtain

$$\begin{aligned} SS_{\text{within}} &= (5 - 4)^2 + (6 - 4)^2 + (4 - 4)^2 \\ &\quad + (5 - 4)^2 + (0 - 4)^2 + (16 - 9)^2 \\ &\quad + (5 - 9)^2 + (9 - 9)^2 + (10 - 9)^2 \\ &\quad + (5 - 9)^2 + (23 - 24)^2 + (30 - 24)^2 \\ &\quad + (20 - 24)^2 + (20 - 24)^2 + (27 - 24)^2 \\ &\quad + (19 - 25)^2 + (35 - 25)^2 + (15 - 25)^2 \\ &\quad + (26 - 25)^2 + (30 - 25)^2 \end{aligned}$$

$$
\begin{aligned}
&= (1)^2 + (2)^2 + (0)^2 + (-4)^2 + (7)^2 \\
&\quad + (-4)^2 + (0)^2 + (1)^2 + (-4)^2 + (-1)^2 \\
&\quad + (6)^2 + (-4)^2 + (-4)^2 + (3)^2 + (-6)^2 \\
&\quad + (10)^2 + (-10)^2 + (1)^2 + (5)^2 \\
&= 1 + 4 + 0 + 1 + 16 + 49 + 16 + 0 + 1 + 16 \\
&\quad + 1 + 36 + 16 + 16 + 9 + 36 + 100 + 100 \\
&\quad + 1 + 25 \\
&= 444
\end{aligned}
$$

Notice that the within-groups sum of squares could have been obtained simply by combining the sum of squares within each group. That is, with four groups,

$$
\begin{aligned}
SS_{\text{within}} &= \Sigma(X_1 - \overline{X}_1)^2 + \Sigma(X_2 - \overline{X}_2)^2 \\
&\quad + \Sigma(X_3 - \overline{X}_3)^2 + \Sigma(X_4 - \overline{X}_4)^2
\end{aligned}
$$

From Table 8.1, we have

$$SS_{\text{within}} = 22 + 82 + 78 + 262 = 444$$

CALCULATING THE BETWEEN-GROUPS SUM OF SQUARES The between-groups sum of squares represents the *sum of the squared deviations of every group mean from the total mean*. Accordingly, we must determine the difference between each group mean and the total mean ($\overline{X}_{\text{group}} - \overline{X}_{\text{total}}$), square this deviation, multiply by the number of scores in that group, and add these quantities. Summing across groups, we obtain the following definitional formula for the between-groups sum of squares:

$$SS_{\text{between}} = \Sigma N_{\text{group}}(\overline{X}_{\text{group}} - \overline{X}_{\text{total}})^2$$

where N_{group} = number of scores in any group

$\overline{X}_{\text{group}}$ = mean of any group

$\overline{X}_{\text{total}}$ = mean of all groups combined

We apply the formula to the data in Table 8.1, and obtain

$$
\begin{aligned}
SS_{\text{between}} &= 5(4 - 15.5)^2 + 5(9 - 15.5)^2 \\
&\quad + 5(24 - 15.5)^2 + 5(25 - 15.5)^2 \\
&= 5(-11.5)^2 + 5(-6.5)^2 + 5(8.5)^2 + 5(9.5)^2 \\
&= 5(132.25) + 5(42.25) + 5(72.25) + 5(90.25) \\
&= 661.25 + 211.25 + 361.25 + 451.25 \\
&= 1{,}685
\end{aligned}
$$

Thus, the sums of squares are

$$
\begin{aligned}
SS_{\text{total}} &= 2{,}129 \\
SS_{\text{within}} &= 444 \\
SS_{\text{between}} &= 1{,}685
\end{aligned}
$$

Notice that the total sum of squares is equal to the within-groups and between-groups sums of squares added together. This relationship among the three sums of squares can be used as a check on your work.

8.2.2: Computing Sums of Squares

The definitional formulas for total, within-groups, and between-groups sums of squares are based on the manipulation of deviation scores, a time-consuming and difficult process. Fortunately, we may instead employ the following much simpler computational formulas to obtain results that are identical (except for rounding errors) to the lengthier definitional formulas:

$$
\begin{aligned}
SS_{\text{total}} &= \Sigma X^2_{\text{total}} - N_{\text{total}}\overline{X}^2_{\text{total}} \\
SS_{\text{within}} &= \Sigma X^2_{\text{total}} - \Sigma N_{\text{group}}\overline{X}^2_{\text{group}} \\
SS_{\text{between}} &= \Sigma N_{\text{group}}\overline{X}^2_{\text{group}} - N_{\text{total}}\overline{X}^2_{\text{total}}
\end{aligned}
$$

where $\Sigma X^2_{\text{total}}$ = all the scores squared and then summed

X_{total} = total mean of all groups combined

N_{total} = total number of scores in all groups combined

$\overline{X}_{\text{group}}$ = mean of any group

N_{group} = number of scores in any group

The raw scores in Table 8.1 have been set up in Table 8.2 for the purpose of illustrating the use of the computational sum-of-squares formulas.

Table 8.2 Computations for Life-Satisfaction Data

Widowed		Divorced	
X_1	X_1^2	X_2	X_2^2
5	25	16	256
6	36	5	25
4	16	9	81
5	25	10	100
0	0	5	25
$\Sigma X_1 = 20$	$\Sigma X_1^2 = 102$	$\Sigma X_2 = 45$	$\Sigma X_2^2 = 487$
$\overline{X}_1 = \frac{20}{5} = 4$		$\overline{X}_2 = \frac{45}{5} = 9$	
Never Married		**Married**	
X_3	X_3^2	X_4	X_4^2
23	529	19	361
30	900	35	1,225
20	400	15	225
20	400	26	676
27	729	30	900
$\Sigma X_3 = 120$	$\Sigma X_3^2 = 2{,}958$	$\Sigma X_4 = 125$	$\Sigma X_4^2 = 3{,}387$
$\overline{X}_3 = \frac{120}{5} = 24$		$\overline{X}_4 = \frac{125}{5} = 25$	
$N_{\text{total}} = 20$	$\overline{X}_{\text{total}} = 15.5$	$\Sigma X_{\text{total}} = 310$	$\Sigma X^2_{\text{total}} = 6{,}934$

Note that before applying the formulas, we must first obtain the sum of scores (ΣX_{total}), sum of squared scores (ΣX^2_{total}), total number of scores (N_{total}), and mean ($\overline{X}_{total}$) for all groups combined:

$$\Sigma X_{total} = \Sigma X_1 + \Sigma X_2 + \Sigma X_3 + \Sigma X_4$$
$$= 20 + 45 + 120 + 125$$
$$= 310$$
$$\Sigma X^2_{total} = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 + \Sigma X_4^2$$
$$= 102 + 487 + 2{,}958 + 3{,}387$$
$$= 6{,}934$$
$$N_{total} = N_1 + N_2 + N_3 + N_4$$
$$= 5 + 5 + 5 + 5$$
$$= 20$$
$$\overline{X}_{total} = \frac{\Sigma X_{total}}{N_{total}}$$
$$= \frac{310}{20}$$
$$= 15.5$$

Next, we move on to calculating the following sums of squares:

$$SS_{total} = \Sigma X^2_{total} - N_{total}\,\overline{X}^2_{total}$$
$$= 6{,}934 - (20)(15.5)^2$$
$$= 6{,}934 - (20)(240.25)$$
$$= 6{,}934 - 4{,}805$$
$$= 2{,}129$$
$$SS_{within} = \Sigma X^2_{total} - \Sigma N_{group}\,\overline{X}^2_{group}$$
$$= 6{,}934 - [(5)(4)^2 + (5)(9)^2 + (5)(24)^2 + (5)(25)^2]$$
$$= 6{,}934 - [(5)(16) + (5)(81) + (5)(576) + (5)(625)]$$
$$= 6{,}934 - (80 + 405 + 2880 + 3125)$$
$$= 6{,}934 - 6{,}490$$
$$= 444$$
$$SS_{between} = \Sigma N_{group}\,\overline{X}^2_{group} - N_{total}\,\overline{X}^2_{total}$$
$$= 6{,}490 - 4{,}805$$
$$= 1{,}685$$

These results agree with the computations obtained using the definitional formulas.

8.3: Mean Square

Objective: Illustrate how the mean square overcomes the problem of sample size affecting the calculation of variation

As we might expect from a measure of variation, the size of the sums of squares tends to become larger as variation increases. For example, SS = 10.9 probably designates greater variation than SS = 1.3. However, the sum of squares also gets larger with increasing sample size, so $N = 200$ will yield a larger SS than $N = 20$. As a result, the sum of squares cannot be regarded as an entirely "pure" measure of variation, unless, of course, we can find a way to control for the number of scores involved.

Fortunately, such a method exists in a measure of variation known as the **mean square** (or *variance*), which we obtain by dividing $SS_{between}$ or SS_{within} by the appropriate degrees of freedom. Recall that earlier in this course, we similarly divided $\sum(X - \overline{X})^2$ by N to obtain the variance. Therefore,

$$MS_{between} = \frac{SS_{between}}{df_{between}}$$

where $MS_{between}$ = between-groups mean square
$SS_{between}$ = between-groups sum of squares
$df_{between}$ = between-groups degrees of freedom

and

$$MS_{within} = \frac{SS_{within}}{df_{within}}$$

where MS_{within} = within-groups mean square
SS_{within} = within-groups sum of squares
df_{within} = within-groups degrees of freedom

But we must still obtain the appropriate degrees of freedom. For the between-groups mean square,

$$df_{between} = k - 1$$

where k = number of groups

To find degrees of freedom for within-groups mean square,

$$df_{within} = N_{total} - k$$

where N_{total} = total number of scores in all groups combined
k = number of groups

Illustrating with the data from Table 8.2, for which $SS_{between} = 1{,}685$ and $SS_{within} = 444$, we calculate our degrees of freedom as follows:

$$df_{between} = 4 - 1$$
$$= 3$$

and

$$df_{within} = 20 - 4$$
$$= 16$$

We are now prepared to obtain the following mean squares:

$$MS_{between} = \frac{1{,}685}{3}$$
$$= 561.67$$

and

$$MS_{within} = \frac{444}{16}$$
$$= 27.75$$

These then are the between and within variances, respectively.

8.4: The *F* Ratio

Objective: Analyze reasons as to why the larger the *F* ratio, the more likely we are to reject the null hypothesis

The analysis of variance yields an *F* ratio in which variation between groups and variation within groups are compared. We are now ready to specify the degree of each type of variation as measured by mean squares. Therefore, the *F* ratio can be regarded as indicating the size of the between-groups mean square relative to the size of the within-groups mean square, or

$$F = \frac{MS_{between}}{MS_{within}}$$

For the data in Table 8.2,

$$F = \frac{561.67}{27.75}$$
$$= 20.24$$

Having obtained an *F* ratio, we must now determine whether it is large enough to reject the null hypothesis and accept the research hypothesis that satisfaction with life differs by marital status. The larger our calculated *F* ratio (the larger the $MS_{between}$ and the smaller the MS_{within}), the more likely we are to obtain a statistically significant result.

But exactly how do we recognize a significant *F* ratio? Recall that earlier in the course, our obtained *t* ratio was compared against a table *t* ratio for the .05 level of significance with the appropriate degrees of freedom. Similarly, we must now interpret our calculated *F* ratio with the aid of Table D in Appendix C. Table D contains a list of critical *F* ratios; these are the *F* ratios that we must obtain to reject the null hypothesis at the .05 and .01 levels of significance. As was the case with the *t* ratio, exactly which *F* value we must obtain depends on its associated degrees of freedom. Therefore, we enter Table D looking for the two df values, between-groups degrees of freedom and within-groups degrees of freedom. Degrees of freedom associated with the numerator ($df_{between}$) have been listed across the top of the page, and degrees of freedom associated with the denominator (df_{within}) have been placed down the left side of the table. The body of Table D presents critical *F* ratios at the .05 and .01 significance levels.

For the data in Table 8.2, we have found $df_{between} = 3$ and $df_{within} = 16$. Thus, we move in Table D to the column marked df = 3 and continue down the page from that point until we arrive at the row marked df = 16. By this procedure, we find that a significant *F* ratio at the $\alpha = 0.5$ level must exceed 3.24, and at the $\alpha = 0.1$ level it must exceed 5.29. Our calculated *F* ratio is 20.24. As a result, we reject the null hypothesis and accept the research hypothesis: Marital status appears to affect life satisfaction.

The results of our analysis of variance can be presented in a summary table such as the one shown in Table 8.3. It has become standard procedure to summarize an analysis of variance in this manner. The total sum of squares ($SS_{total} = 2{,}129$) is decomposed into two parts: the between-groups sum of squares ($SS_{between} = 1{,}685$) and the within-groups sum of squares ($SS_{within} = 444$). Each source of sum of squares is converted to mean square by dividing by the respective degrees of freedom. Finally, the *F* ratio (mean square *between* divided by mean square *within*) is calculated, which can be compared to the table critical value to determine significance.

Table 8.3 Analysis of Variance Summary Table for the Data in Table 8.2

Source of Variation	SS	df	MS	*F*
Between groups	1,685	3	561.67	20.24
Within groups	444	16	27.75	
Total	2,129	19		

8.4.1: Two Contrasting Examples of ANOVA

To review some of the concepts presented thus far, consider Table 8.4, which shows two contrasting situations. Both Case 1 and Case 2 consist of three samples (A, B, and C) with sample means $\overline{X}_A = 3$, $\overline{X}_B = 7$, and $\overline{X}_C = 11$ and with $N = 3$ in each sample. Because the means are the same in both data sets, the between-groups sums of squares are identical ($SS_{between} = 96$).

Table 8.4 Two Examples of Analysis of Variance

Case 1 Data			Analysis-of-Variance Summary Table				
Sample A	*Sample B*	*Sample C*	*Source of Variation*	*SS*	*df*	*MS*	*F*
2	6	10	Between groups	96	2	48	48
3	7	11	Within groups	6	6	1	
4	8	12					
Mean 3	7	11	Total	102	8		

Distributions

	A	A	A		B	B	B		C	C	C			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Case 2 Data			Analysis-of-Variance Summary Table				
Sample A	*Sample B*	*Sample C*	*Source of Variation*	*SS*	*df*	*MS*	*F*
1	3	7	Between groups	96	2	48	3.2
3	5	12	Within groups	90	6	15	
5	13	14					
Mean 3	7	11	Total	186	8		

Distributions

		B		B										
A		A		A		C					C	B	C	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

In Case 1, the three samples are clearly different. It would seem then that we should be able to infer that the population means are different. Relative to between-groups variation (the differences between the sample means), the within-groups variation is rather small. Indeed, there is as much as a 48-to-1 ratio of between-groups mean square to within-groups mean square. Thus, $F = 48$ and is significant. Although the sample means and between-groups sum of squares are the same for Case 2, there is far more dispersion within groups, causing the samples to overlap quite a bit. The samples hardly appear as distinct as in Case 1, and so it would seem unlikely that we could generalize the differences between the sample means to differences between population means. The within-groups mean square is 15. Therefore, the ratio of between-to-within mean square is only 48-to-15, yielding a nonsignificant *F* ratio of 3.2.

SUMS OF SQUARES, MEAN SQUARES, AND THE *F* RATIO Before moving to a step-by-step illustration, it will help to review the relationship between sums of squares, mean squares, and the *F* ratio. SS_{within} represents the sum of squared deviations of scores from their respective sample means, while the $SS_{between}$ involves the sum of squared deviations of the sample means from the total mean for all groups combined. The mean square is the sum of squares, within or between, divided by the corresponding degrees of freedom. Earlier, we estimated the variance of a population using sample data through dividing the sum of squared deviations from the sample mean by the degrees of freedom (for one sample, $N - 1$). Now, we are confronted with two sources of variance, within and between, for each of which we divide the sum of squared deviations (within or between) by the appropriate degrees of freedom to obtain estimates of the variance (within or between) that exists in the population.

Finally, the *F* statistic is the ratio of two variances, for example, between-groups variance (estimated by $MS_{between}$) divided by within-groups variance (estimated by MS_{within}). Even if the population means for all the groups were equal—as stipulated by the null hypothesis—the sample means would not necessarily be identical because of sampling error. Thus, even if the null hypothesis is true, $MS_{between}$ would still reflect sampling variability, but only sampling variability. On the other hand, MS_{within} reflects only sampling variability, whether the null hypothesis is true or false. Therefore, if the null hypothesis of no population mean differences holds, the *F* ratio should be approximately equal to 1, reflecting the ratio of two estimates of sampling variability. But, as the group means diverge, the numerator of the *F* ratio will grow, pushing the *F* ratio over 1. The *F* table then indicates how large an *F* ratio we require—that is, how many times larger $MS_{between}$ must be in comparison to MS_{within}, to reject the null hypothesis.

To gain more understanding of analysis of variance, review the following step-by-step illustration.

Step-by-Step Illustration: Analysis of Variance

To provide a step-by-step illustration of an analysis of variance, suppose that a social researcher interested in employment discrimination issues conducts a simple experiment to assess whether male law school graduates are favored over females by large and well-established law firms. She asks the hiring partners at 15 major firms to rate a set of resumes in terms of background and potential on a scale of 0 to 10, with higher scores indicating stronger interest in the applicant. The set of resumes given to the 15 respondents are identical with one exception. On one of the resumes, the name is changed to reflect a male applicant (applicant's name is "Jeremy Miller"), a gender-ambiguous applicant (applicant's name is "Jordan Miller"), or a female applicant (applicant's name is "Janice Miller"); everything else about the resume is the same in all conditions. The 15 respondents are randomly assigned one of these three resumes within their packets of resumes. The other resumes in the packets are the same for all of the hiring partners. These other resumes are included to provide a better context for the experiment, yet only the ratings of the Jeremy/Jordan/Janice Miller resumes are of interest.

The researcher establishes these hypotheses:

Null hypothesis: ($\mu_1 = \mu_2 = \mu_3$)	*Ratings of applicants do not differ based on gender*
Research hypothesis: (some $\mu_i \neq \mu_j$)	*Ratings of applicants differ based on gender*

The rating scores given by the 15 hiring partners who had been randomly assigned to the three groups based on the apparent gender of the key resume are as follows:

Male ($N_1 = 5$)		**Gender-Neutral ($N_2 = 5$)**		**Female ($N_3 = 5$)**	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
6	36	2	4	3	9
7	49	5	25	2	4
8	64	4	16	4	16
6	36	3	9	4	16
4	16	5	25	3	9
$\Sigma X_1 = 31$	$\Sigma X_1^2 = 201$	$\Sigma X_2 = 19$	$\Sigma X_2^2 = 79$	$\Sigma X_3 = 16$	$\Sigma X_3^2 = 54$

Step 1 Find the mean for each sample.

$$\bar{X}_1 = \frac{\Sigma X_1}{N_1} = \frac{31}{5} = 6.2$$

$$\bar{X}_2 = \frac{\Sigma X_2}{N_2} = \frac{19}{5} = 3.8$$

$$\bar{X}_3 = \frac{\Sigma X_3}{N_3} = \frac{16}{5} = 3.2$$

Notice that differences do exist, the tendency being for the apparently male applicant to be rated higher than the gender-neutral and apparently female applicant.

Step 2 Find the sum of scores, sum of squared scores, total number of subjects, and mean for all groups combined.

$$\Sigma X_{total} = \Sigma X_1 + \Sigma X_2 + \Sigma X_3 = 31 + 19 + 16 = 66$$

$$\Sigma X^2_{total} = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 = 201 + 79 + 54 = 334$$

$$N_{total} = N_1 + N_2 + N_3 = 5 + 5 + 5 = 15$$

$$\bar{X}_{total} = \frac{\Sigma X_{total}}{N_{total}} = \frac{66}{15} = 4.4$$

Step 3 Find the total sum of squares.

$$SS_{total} = \Sigma X^2_{total} - N_{total}\bar{X}^2_{total} = 334 - (15)(4.4)^2 = 334 - (15)(19.36) = 334 - 290.4 = 43.6$$

Step 4 Find the between-groups sum of squares.

$$SS_{between} = \Sigma N_{group}\bar{X}^2_{group} - N_{total}\bar{X}^2_{total}$$
$$= [(5)(6.2)^2 + (5)(3.8)^2 + (5)(3.2)^2] - (15)(4.4)^2$$
$$= [(5)(38.44) + (5)(14.44) + (5)(10.24)] - (15)(19.36)$$
$$= (192.2 + 72.2 + 51.2) - 290.4$$
$$= 315.6 - 290.4$$
$$= 25.2$$

Step 5 Find the within-groups sum of squares.

$$SS_{within} = \Sigma X^2_{total} - \Sigma N_{group}\bar{X}^2_{group}$$
$$= 334 - [(5)(6.2)^2 + (5)(3.8)^2 + (5)(3.2)^2]$$
$$= 334 - [(5)(38.44) + (5)(14.44) + (5)(10.24)]$$
$$= 334 - (192.2 + 72.2 + 51.2)$$
$$= 334 - 315.6$$
$$= 18.4$$

Step 6 Find the between-groups degrees of freedom.

$$df_{between} = k - 1 = 3 - 1 = 2$$

Step 7 Find the within-groups degrees of freedom.

$$df_{within} = N_{total} - k = 15 - 3 = 12$$

Step 8 Find the between-groups mean square.

$$MS_{between} = \frac{SS_{between}}{df_{between}} = \frac{25.2}{2} = 12.6$$

Step 9 Find the within-groups mean square.

$$MS_{within} = \frac{SS_{within}}{df_{within}} = \frac{18.4}{12} = 1.53$$

Step 10 Obtain the F ratio.

$$F = \frac{MS_{between}}{MS_{within}} = \frac{12.6}{1.53} = 8.24$$

Step 11 Compare the obtained F ratio with the appropriate value found in Table D.

obtained F ratio $= 8.24$

table F ratio $= 3.88$

$$df = 2 \text{ and } 12$$
$$\alpha = .05$$

As shown in Step 11, to reject the null hypothesis at the .05 significance level with 2 and 12 degrees of freedom, our calculated *F* ratio must exceed 3.88. Because we have obtained an *F* ratio of 8.24, we can reject the null hypothesis and accept the research hypothesis. Specifically, we conclude that the ratings given law firm applicants tend to differ on the basis of their apparent gender.

8.5: A Multiple Comparison of Means

Objective: **Calculate Tukey's HSD as a useful test for investigating the multiple comparison of means**

A significant *F* ratio informs us of an overall difference among the groups being studied. If we were investigating a difference between only two sample means, no additional analysis would be needed to interpret our result: In such a case, either the obtained difference is statistically significant or it is not. However, when we find a significant *F* for the differences among three or more means, it may be important to determine exactly where the significant differences lie. For example, in the foregoing step-by-step illustration, the social researcher uncovered statistically significant differences in the ratings of law firm applicants based on their gender.

Consider the possibilities raised by this significant *F* ratio: The ratings given the male applicant may differ significantly from those given the gender-neutral applicant, the ratings given the male applicant may differ significantly from those given the female applicant; and the ratings given the gender-neutral applicant may differ significantly from those given the female applicant. As explained earlier in this chapter, obtaining a *t* ratio for each comparison—$\overline{X}_1$ versus $\overline{X}_2$, $\overline{X}_2$ versus $\overline{X}_3$, and $\overline{X}_1$ versus $\overline{X}_3$ —would entail a good deal of work and, more importantly, would increase the probability of Type I error. Fortunately, statisticians have developed a number of other tests for making multiple comparisons after obtaining a significant *F* ratio to pinpoint where the significant mean differences lie. We introduce **Tukey's HSD** (Honestly Significant Difference), one of the most useful tests for investigating the multiple comparison of means. Tukey's HSD is used only after a significant *F* ratio has been obtained. By Tukey's method, we compare the difference between any two mean scores against HSD. A mean difference is statistically significant only if it exceeds HSD. By formula,

$$HSD = q\sqrt{\frac{MS_{within}}{N_{group}}}$$

where
q = value in Table E of Appendix C at a given level of significance for the total number of group means being compared
MS_{within} = within-groups mean square (obtained from the analysis of variance)
N_{group} = number of subjects in each group (assumes the same number in each group)[1]

Unlike the *t* ratio, HSD takes into account that the likelihood of Type I error increases as the number of means being compared increases. The *q* value depends upon the number of group means, and the larger the number of group means, the more conservative HSD becomes with regard to rejecting the null hypothesis. As a result, fewer significant differences will be obtained with HSD than with the *t* ratio. Moreover, a mean difference is more likely to be significant in a multiple comparison of three means than in a multiple comparison of four or five means.

8.5.1: Using Tukey's HSD

Now, we will go through the step-by-step procedure to analyze the use of HSD in multiple comparison of three means.

To illustrate the use of HSD, let us return to the previous example in which ratings of law school graduates by large law firm hiring partners differed significantly based on the apparent gender of the applicant. More specifically, a significant *F* ratio was obtained for differences among the mean ratings for the three groups of respondents—6.2 for those evaluating a resume from a male applicant, 3.8 for those assessing a gender neutral applicant, and 3.2 for those evaluating a resume from a female applicant, respectively. That is,

$$\overline{X}_1 = 6.2$$
$$\overline{X}_2 = 3.8$$
$$\overline{X}_3 = 3.2$$

Step-by-Step Illustration: HSD for Analysis of Variance

Step 1 Construct a table of differences between ordered means.

Arrange the means from smallest to largest: 3.2, 3.8, and 6.2. These mean scores are placed in a table so that the difference between each pair of

[1]Tukey's method can be used for comparison of groups of unequal size. In such cases, N_{group} is replaced by what is called the harmonic mean of the group sizes. The harmonic mean of the group sizes is the reciprocal of the mean of the reciprocal group sizes. That is,

$$N_{harmonic} = \frac{k}{\frac{1}{N_1} + \frac{1}{N_2} + \cdots + \frac{1}{N_k}}$$ where k is the number of groups being compared, and $N_1, N_2, \ldots, N_k$ are the respective group sizes.

means is shown in a matrix. Thus, the difference between $\bar{X}_1$ and $\bar{X}_2$ is .6; the difference between $\bar{X}_1$ and $\bar{X}_3$ is 3.0; and the difference between $\bar{X}_2$ and $\bar{X}_3$ is 2.4. The subscripts for the group means should not change when arranged in order. Thus, for example, $\bar{X}_2$ represents the mean of the group originally designated as number 2, not the second-highest group mean.

	$\bar{X}_3 = 3.2$	$\bar{X}_2 = 3.8$	$\bar{X}_1 = 6.2$
$\bar{X}_3 = 3.2$	—	.6	3.0
$\bar{X}_2 = 3.8$	—	—	2.4
$\bar{X}_1 = 6.2$	—	—	—

Step 2 Find q in Table E in Appendix C.

To find q from Table E, we must have (1) the degrees of freedom (df) for MS_{within}, (2) the number of group means k, and (3) a significance level of either .01 or .05. We already know from the analysis of variance that $df_{within} = 12$. Therefore, we look down the left-hand column of Table E until we arrive at 12 degrees of freedom. Second, because we are comparing three mean scores, we move across Table E to a number of group means (k) equal to three. Assuming a .05 level of significance, we find that $q = 3.77$.

Step 3 Find HSD.

$$\begin{aligned} HSD &= q\sqrt{\frac{MS_{within}}{N_{group}}} \\ &= 3.77\sqrt{\frac{1.53}{5}} \\ &= 3.77\sqrt{.306} \\ &= (3.77)(.553) \\ &= 2.08 \end{aligned}$$

Step 4 Compare the HSD against the table of differences between means.

To be regarded as statistically significant, any obtained difference between means must exceed the HSD (2.08). Referring to the table of differences between means, we find that the mean rating difference of 3.0 between $\bar{X}_3$ (female applicant) and $\bar{X}_1$ (male applicant) and mean difference of 2.4 between $\bar{X}_1$ (male applicant) and $\bar{X}_2$ (gender-neutral applicant) are greater than HSD = 2.08. As a result, we conclude that these differences between means are statistically significant at the .05 level. Finally, the difference in mean rating of .6 between $\bar{X}_3$ (female applicant) and $\bar{X}_2$ (gender-neutral applicant) is not significant because it is less than HSD.

8.6: Two-Way Analysis of Variance

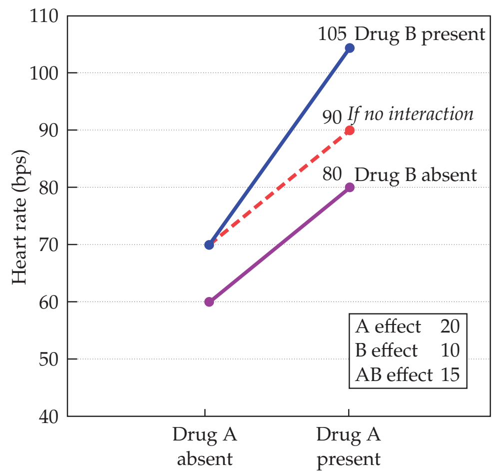
Objective: Examine the influence of two factors or independent variables using two-way analysis of variance

Analysis of variance can be used to examine the effect of two or more factors (or variables) simultaneously. Suppose that a researcher is interested in the effect of particular drugs on heart rate. He first takes his subject's pulse and records the result as 60 beats per minute (bpm). He then administers Drug A to his subject and observes that the heart rate increases to 80 bpm. Thus, the effect produced by Drug A is 20 bpm (80 − 60 = 20). After waiting until the effect of Drug A wears off (until the heart rate returns to 60), the researcher then administers Drug B, observing that the heart rate increases to 70 bpm. Thus, the effect of Drug B is 10 bpm (70 − 60 = 10). Finally, the researcher is curious about what the result would be if he were to give his subject both drugs at the same time. The effect of Drug A is 20 and the effect of Drug B is 10. Would the combined effect of Drugs A and B be additive—that is, would the heart rate increase by 20 + 10 = 30 bpm, reaching 90 bpm upon administration of both? Testing this out, the researcher observes his subject's heart rate actually skyrockets to 105 bpm, 15 bpm more than would have been expected by adding the separate effects of the drugs.

This simple example illustrates **main effects** and **interaction effects**. A main effect is the influence on heart rate of either Drug A or Drug B by itself. Here, Drug A has

Figure 8.3 Effects of Drugs A and B on Heart Rate

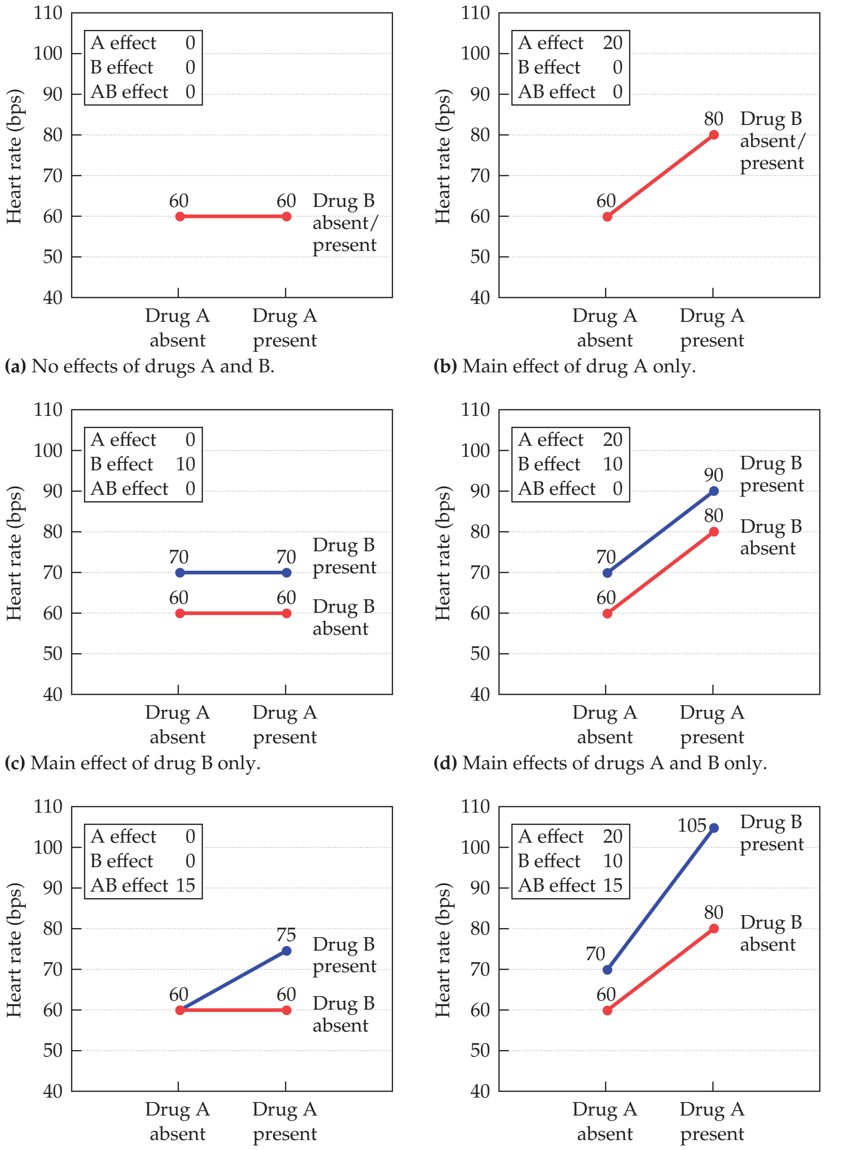
The main and interaction effects are illustrated here. The heart rate is calibrated on the vertical axis, the absence or presence of Drug A on the horizontal axis, and the absence and presence of Drug B is shown on separate lines.



a main effect of 20 and Drug B has a main effect of 10. The interaction effect refers to the extra impact of the two drugs together, beyond the sum of their separate or main effects. Therefore, the interaction effect of Drugs A and B is 15. The main and interaction effects are illustrated in Figure 8.3. The heart rate is calibrated on the vertical axis, the absence or presence of Drug A is shown on the horizontal axis, and separate lines are employed for the absence and presence of Drug B. Had the effects of the two drugs been additive (noninteractive), then the lines would have been parallel, as shown.

There are many different possibilities for main and interaction effects, as illustrated in Figure 8.4. Graph (a) shows what happens when there are no effects of any kind. Under all four conditions (no drugs, Drug A alone, Drug B alone, and both drugs together), the heart rate is 60. Graph (b) shows what happens when only Drug A has an effect on heart rate. Here, the heart rate increases to 80, whether or not Drug B is present. Graph (c) shows what happens when only Drug B has an effect on heart rate. Here, the heart rate increases to 70, whether or not Drug A is present. Graph (d) shows what happens when both drugs have main effects but do not interact. Drug A increases the heart rate by 20, regardless of the presence of B, and Drug B increases the heart rate by 10, regardless of the presence of A. Together they increase the heart rate by 30, which is the

Figure 8.4 Main and Interaction Effects



(a) No effects of drugs A and B.

(b) Main effect of drug A only.

(c) Main effect of drug B only.

(d) Main effects of drugs A and B only.

(e) Interaction effects of drugs A and B only.

(f) Main and interaction effects of drugs A and B.

sum total of the main effects. Graph (e) illustrates the situation that occurs when there is only an interaction but no main effects. By themselves the drugs have no effect, but together they cause the heart rate to increase to 75, an interaction effect of 15. Finally, graph (f) shows the presence of all three effects; a main effect for A of 20, a main effect for B of 10, and an interaction effect (AB) of 15.

The illustration in Figure 8.4 is designed to introduce the logic of determining main and interaction effects.

The mechanics of two-way analysis of variance depart from our hypothetical example in two important respects:

- First, the main and interaction effects shown in the figure were derived by comparing particular treatment combinations against the control condition (no drugs) as a baseline. In the analysis of variance formulas to follow, treatment combinations are compared against the total mean of all groups as a baseline,
- Second, in actual research settings, a researcher would obviously observe more than one subject to have confidence in the results. In addition, it is often necessary to make comparisons between different groups of subjects—for example, a group that gets Drug A only, a group that gets Drug B only, a group that gets both Drugs A and B, and a control.

The same way of thinking about main and interaction effects applies to groups of subjects, except that we focus on differences between group means. For example, we would analyze the mean heart rate for the group given only Drug A, for those given only Drug B, for those given both Drugs A and B, and for those given neither drug. In this chapter we only consider the situation of independent groups—different subjects for different conditions. More advanced texts will extend this approach to repeated measures designs—the same subjects across different conditions.

8.6.1: The Influence of Two Factors in the Same Experiment

As we saw earlier in the chapter, analysis of variance can be useful for examining group differences when three or more sample means are being compared. Actually, the analysis of variance discussed earlier is the simplest kind and is called a one-way analysis of variance, because it represents the effect of different categories or levels of a single factor or independent variable on a dependent variable. When we examine the influence of two factors or independent variables together in the same experiment (such as with Drugs A and B), it is called two-way analysis of variance.

Let's consider a simple, yet far more realistic example. Suppose a social researcher specializing in media effects wants to determine if sex and violence in films impact the perceptions and attitudes of viewers. After gaining informed consent from 16 students from his undergraduate media class, he randomly divides his volunteers into four groups with four students each. One group watches a series of nonviolent, nonsexual films; another a series of violent, nonsexual films, another a series of nonviolent, sexual films; and the last group a series of violent, sexual films. Following the study period, the researcher/instructor administers to all 16 volunteers a test to measure empathy for victims of rape, with a scale ranging from 0 for no empathy to 10 for a high level of empathy. The results are shown in Table 8.5.

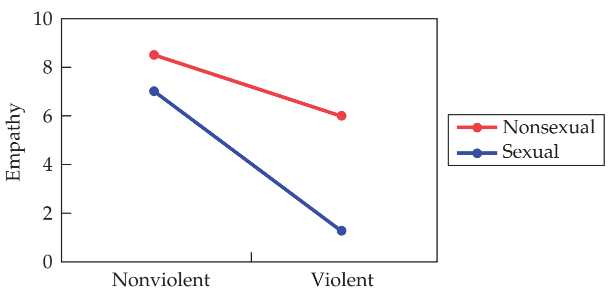
Table 8.5 Empathy Scores by Film Content Combination

Nonviolent		Violent	
Nonsexual	*Sexual*	*Nonsexual*	*Sexual*
8	9	6	2
10	5	4	1
7	7	8	1
9	7	6	2

One might be tempted to analyze these data using a one-way analysis of variance with four groups. However, this would not capture the fact that one of the groups combines the attributes of two others. That is, the group that watches sexually violent films is exposed to the effect of violent content, sexual content, and their interaction. Two-way analysis of variance permits us to disentangle main and interaction effects appropriately.

Figure 8.5 displays the group means to assess the main effects of violent and sexual content and their interaction on empathy levels. Not only do both content types appear to lower empathy levels of viewers, but the combination reduces empathy even more. Of course, as with any of the tests of differences covered thus far in the text, we still need to determine if the observed differences between the group means—and thus the main and interaction effects—are greater than one would obtain by chance. As with one-way analysis of variance, this involves decomposing the total sum of squares into between and within sources. But here, we shall further divide the between sum of squares into portions resulting from the effect of violent content (A), the effect of sexual content (B), and the interaction effect of violent and sexual content (AB).

Figure 8.5 Plot of Group Means for Film Experiment



8.6.2: Dividing the Total Sum of Squares

In our presentation of one-way analysis of variance, we learned that the sum of squares refers to the sum of squared deviations from a mean. For a one-way analysis of variance, the total sum of squares (SS_{total}) was divided into two components, the between-groups sum of squares ($SS_{between}$), representing variation between group means as a result of an independent variable, and the within-groups sum of squares (SS_{within}), representing random variation among the scores of the members of the same group. The total sum of squares was shown to be equal to a combination of its within- and between-groups components. Thus,

$$SS_{total} = SS_{between} + SS_{within}$$

In the case of a two-way analysis of variance, the total sum of squares can again be divided into its within-groups and between-groups components. This time, however, because more than one independent variable is involved, the between-groups sum of squares can be further broken down into

$$SS_{between} = SS_A + SS_B + SS_{AB}$$

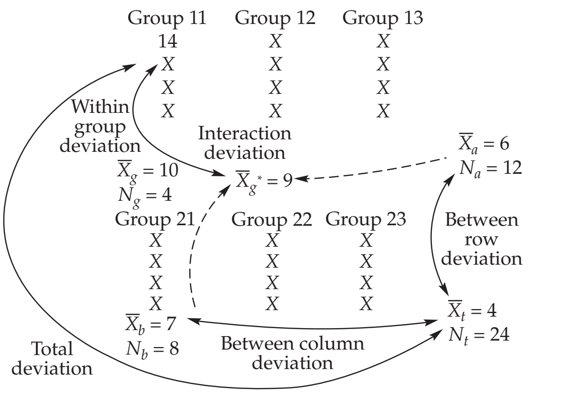
where SS_A is the sum of squares for the main effect A based on variation between levels of factor A

SS_B is the sum of squares for the main effect B based on variation between levels of factor B

SS_{AB} is the sum of squares for the interaction effect of A and B, based on variation between combinations of A and B

AN ILLUSTRATION OF TWO-WAY ANOVA Consider the hypothetical experimental results shown in Figure 8.6. Note that to help us focus on the concepts of total, within-groups, main effect, and interaction sums of squares, only part of the data set is shown. For the sake of brevity, *g* represents the group, *a* is the level of the factor aligned in rows, *b* is the level of the factor aligned in columns, and *t* represents the total.

Figure 8.6 Illustration of Two-Way Analysis of Variance (*g* = Group, *a* = Row Level, *b* = Column Level, *t* = Total)



The score received by the first subject in Group 11 (shown above as 14) is substantially higher than the total mean for the study ($\bar{X}_{total} = 4$). His deviation from the total mean is ($X - \bar{X}_{total} = 10$). Part of his elevated score represents the fact that the groups in which his score is located (his level of Factor A) performed better than average; that is, $(\bar{X}_a - \bar{X}_{total}) = 2$. Part of his elevated score also represents the fact that the groups in which his score is located (his level of Factor B) performed better than average; that is, $(\bar{X}_b - \bar{X}_{total}) = 3$. Note that combining the difference due to Factor A (2) and Factor B (3) would suggest a group mean that is $2 + 3 = 5$ units above the total mean, that is, $\bar{X}^*_{group} = 4 + 5 = 9$. (The notation $\bar{X}^*_{group}$ is used to represent what the group mean would be in the absence of any interaction between Factors A and B.) The deviation of the group mean ($\bar{X}_{group}$) from that produced by adding the two main effects is a result of interaction, $\bar{X}_{group} - \bar{X}_a - \bar{X}_b + \bar{X}_{total} = 1$. Finally, after accounting for deviations or differences due to main and interaction effects, the subject's score remains higher than the group mean. Within the group, his deviation is $X - \bar{X}_{group} = 4$.

All these deviations—the deviations of subjects from the total mean, deviations of subjects from their group means, deviations between Factor A mean and the total mean, deviations between Factor B mean and the total mean, and deviations between group means and those expected by combining main effects, can be squared, and then summed to obtain SS_{total}, SS_{within}, SS_A, SS_B, and SS_{AB}. Specifically,

$$SS_{total} = \Sigma(X - \bar{X}_{total})^2$$

$$SS_{within} = \Sigma(X - \bar{X}_{group})^2$$

$$SS_A = \Sigma N_a(\bar{X}_a - \bar{X}_{total})^2$$

$$SS_B = \Sigma N_b(\bar{X}_b - \bar{X}_{total})^2$$

$$SS_{AB} = \Sigma N_{group}(\bar{X}_{group} - \bar{X}_a - \bar{X}_b + \bar{X}_{total})^2$$

where $\bar{X}_{total}$ = the total mean of all groups combined

$\bar{X}_{group}$ = the mean of any group

$\bar{X}_a$ = the mean of any level of Factor A

$\bar{X}_b$ = the mean of any level of Factor B

N_{group} = the number of cases in any group

N_a = the number of cases in any level of Factor A

N_b = the number of cases in any level of Factor B

CALCULATING THE SUMS OF SQUARES As in the case of one-way analysis of variance, the sums of squares are generally more easily obtained using computational formulas based on raw scores rather than the previous definitional formulas based on deviations. For two-way analysis of variance, the computational formulas for calculating total, within, main effect, and interaction sums of squares are as follows:

$$SS_{total} = \Sigma X^2_{total} - N_{total}\bar{X}^2_{total}$$

$$SS_{within} = \Sigma X^2_{total} - \Sigma N_{group}\bar{X}^2_{group}$$

$$SS_A = \Sigma N_a \bar{X}^2_a - N_{total}\bar{X}^2_{total}$$

$$SS_B = \Sigma N_b \bar{X}^2_b - N_{total}\bar{X}^2_{total}$$

$$SS_{AB} = \Sigma N_{group}\bar{X}^2_{group} - \Sigma N_a \bar{X}^2_a - \Sigma N_b \bar{X}^2_b + N_{total}\bar{X}^2_{total}$$

As with one-way analysis of variance, each of the sums of square is associated with degrees of freedom, which, in turn, are used to derive mean squares. Specifically,

$$df_{within} = N_{total} - ab$$

$$MS_{within} = \frac{SS_{within}}{df_{within}}$$

$$df_A = a - 1$$

$$MS_A = \frac{SS_A}{df_A}$$

$$df_B = b - 1$$

$$MS_B = \frac{SS_B}{df_B}$$

$$df_{AB} = (a - 1)(b - 1)$$

$$MS_B = \frac{SS_{AB}}{df_{AB}}$$

where a is the number of levels of Factor A, and b is the number of levels of Factor B.

Finally, as in the one-way case of analysis of variance, the mean squares for the main and interaction effects are tested for statistical significance by dividing each by the MS_{within}. This produces an F ratio for each effect, which can in turn be compared to the value found in Table D in Appendix C with the appropriate degrees of freedom. That is,

$$F_A = \frac{MS_A}{MS_{within}} \text{ with } a - 1 \text{ and } N_{total} - ab \text{ degrees of freedom}$$

$$F_B = \frac{MS_B}{MS_{within}} \text{ with } b - 1 \text{ and } N_{total} - ab \text{ degrees of freedom}$$

$$F_{AB} = \frac{MS_{AB}}{MS_{within}} \text{ with } (a - 1)(b - 1) \text{ and } N_{total} - ab \text{ degrees of freedom}$$

If any of the F ratios exceed the value in Table D in Appendix C, we can conclude that the corresponding main or interaction effects are statistically significant, that is, larger than one would expect by chance alone. Now, let's review two-way analysis of variance with a step-by-step illustration.

Step-by-Step Illustration: Two-Way Analysis of Variance

As a step-by-step illustration of two-way analysis of variance, let's return to the small experiment on the effects of violence and sex in films on the empathy of viewers toward rape victims. We have four groups of four subjects each, the results of which are displayed as follows in the form of a 2 × 2 table with various sums and means calculated for each group, for each level of Factor A (violent content) aligned in the rows, for each level of Factor B (sexual content) aligned in the columns, and for the entire study.

		Sexual Content (B)				**Total**
		Nonsexual		*Sexual*		
		X	X^2	X	X^2	$N = 8$ $\bar{X} = 7.75$
		8	64	9	81	
		10	100	5	25	
	Nonviolent	7	49	7	49	
		9	81	7	49	
		34	294	28	204	
Violent Content (A)		$N = 4$	$\bar{X} = 8.5$	$N = 4$	$\bar{X} = 7.0$	
		X	X^2	X	X^2	$N = 8$ $\bar{X} = 3.75$
		6	36	2	4	
		4	16	1	1	
	Violent	8	64	1	1	
		6	36	2	4	
		24	152	6	10	
		$N = 4$	$\bar{X} = 6.0$	$N = 4$	$\bar{X} = 1.5$	
	Total	$N = 8$	$\bar{X} = 7.25$	$N = 8$	$\bar{X} = 4.25$	$N = 16$ $\bar{X} = 5.75$

Our objective is to test the null hypothesis that all groups are equal in terms of their population means with the research hypothesis being that not all population means are equal. That is,

Null hypothesis: ($\mu_1 = \mu_2 = \mu_3 = \mu_4$)	*Violence and sex in films have no effect on viewer empathy*
Research hypothesis: (some $\mu_i \neq \mu_j$)	*Violence and sex in films have an effect on viewer empathy*

Step 1 Find the mean for each group.

$$\bar{X}_{\text{group}} = \frac{\Sigma X_{\text{group}}}{N_{\text{group}}}$$

$$= \frac{34}{4} = 8.5 \text{ (nonviolent/nonsexual)}$$

$$= \frac{24}{4} = 6.0 \text{ (violent/nonsexual)}$$

$$= \frac{28}{4} = 7.0 \text{ (nonviolent/sexual)}$$

$$= \frac{6}{4} = 1.5 \text{ (violent/sexual)}$$

Step 2 Find the total mean.

$$\bar{X}_{\text{total}} = \frac{\Sigma X_{\text{total}}}{N_{\text{total}}} = \frac{96}{16} = 5.75$$

Step 3 Find the mean for each level of Factor A.

$$\bar{X}_{\text{A}} = \frac{\Sigma X_{\text{A}}}{N_{\text{A}}}$$

$$= \frac{62}{8} = 7.75 \text{ (nonviolent)}$$

$$= \frac{30}{8} = 3.75 \text{ (violent)}$$

Step 4 Find the mean for each level of Factor B.

$$\bar{X}_{\text{B}} = \frac{\Sigma X_{\text{B}}}{N_{\text{B}}}$$

$$= \frac{58}{8} = 7.25 \text{ (nonviolent)}$$

$$= \frac{34}{8} = 4.25 \text{ (violent)}$$

Step 5 Calculate the total, within, main effect, and interaction sums of squares.

$$\begin{aligned} SS_{\text{total}} &= \Sigma X^2_{\text{total}} - N_{\text{total}}\bar{X}^2_{\text{total}} \\ &= (294 + 204 + 152 + 10) - 16(5.75)^2 \\ &= 660 - 529 \\ &= 131 \end{aligned}$$

$$\begin{aligned} SS_{\text{within}} &= \Sigma X^2_{\text{total}} - \Sigma N_{\text{group}}\bar{X}^2_{\text{group}} \\ &= (294 + 204 + 152 + 10) - [4(8.5)^2 \\ &\quad + 4(6.0)^2 + 4(7.0)^2 + 4(1.5)^2] \\ &= 660 - (289 + 144 + 196 + 9) \\ &= 660 - 638 \\ &= 22 \end{aligned}$$

$$\begin{aligned} SS_{\text{A}} &= \Sigma N_a\bar{X}^2_a - N_{\text{total}}\bar{X}^2_{\text{total}} \\ &= [8(7.75)^2 + 8(3.75)^2] - 16(5.75)^2 \\ &= (480.5 + 112.5) - 529 \\ &= 593 - 529 \\ &= 64 \end{aligned}$$

$$\begin{aligned} SS_{\text{B}} &= \Sigma N_b\bar{X}^2_b - N_{\text{total}}\bar{X}^2_{\text{total}} \\ &= [8(7.25)^2 + 8(4.25)^2] - 16(5.75)^2 \\ &= (420.5 + 144.5) - 529 \\ &= 565 - 529 \\ &= 36 \end{aligned}$$

$$\begin{aligned} SS_{\text{AB}} &= \Sigma N_{\text{group}}\bar{X}^2_{\text{group}} - \Sigma N_a\bar{X}^2_a - \Sigma N_b\bar{X}^2_b + N_{\text{total}}\bar{X}^2_{\text{total}} \\ &= [4(8.5)^2 + 4(6.0)^2 + 4(7.0)^2 + 4(1.5)^2] \\ &\quad - [8(7.75)^2 + 8(3.75)^2] - [8(7.25)^2 \\ &\quad + 8(4.25)^2] + 16(5.75)^2 \\ &= 638 - 593 - 565 - 529 \\ &= 9 \end{aligned}$$

Step 6 Find the degrees of freedom for within, main effects, and interaction sums of squares.

$$\begin{aligned} df_{\text{within}} &= N_{\text{total}} - ab = 16 - 2(2) = 12 \\ df_{\text{A}} &= a - 1 = 2 - 1 = 1 \\ df_{\text{B}} &= b - 1 = 2 - 1 = 1 \\ df_{\text{AB}} &= (a - 1)(b - 1) = (2 - 1)(2 - 1) = 1 \end{aligned}$$

Step 7 Find the mean squares for within, main effects, and interaction sums of squares.

$$MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{22}{12} = 1.833$$

$$MS_{\text{A}} = \frac{SS_{\text{A}}}{df_{\text{A}}} = \frac{64}{1} = 64$$

$$MS_{\text{B}} = \frac{SS_{\text{B}}}{df_{\text{B}}} = \frac{36}{1} = 36$$

$$MS_{\text{AB}} = \frac{SS_{\text{AB}}}{df_{\text{AB}}} = \frac{9}{1} = 9$$

Step 8 Obtain the *F* ratios for the main effects and the interaction effect.

$$F_{\text{A}} = \frac{MS_{\text{A}}}{MS_{\text{within}}} = \frac{64}{1.833} = 34.909$$

$$F_{\text{B}} = \frac{MS_{\text{B}}}{MS_{\text{within}}} = \frac{36}{1.833} = 19.636$$

$$F_{\text{AB}} = \frac{MS_{\text{AB}}}{MS_{\text{within}}} = \frac{9}{1.833} = 4.909$$

Step 9 Compare the *F* ratios with the appropriate value found in Table D in Appendix C.

All three *F* ratios have 1 and 12 degrees of freedom and exceed 4.75, the *F* ratio from Table D for the .05 level of significance.

Step 10 Arrange the results in an analysis of variance summary table.

Source	SS	df	MS	*F*
Violent content (A)	64	1	64.0	34.909
Sexual content (B)	36	1	36.0	19.636
Interaction (AB)	9	1	9.0	4.909
Within group	22	12	1.833	
Total	131	15		

Based on these results, we can reject the null hypothesis that all group means are equal in the populations. Furthermore, we find that there are significant main effects of violent and sexual content as well as a significant interaction. In other words, not only do both forms of content tend to reduce viewer empathy, but in combination the effect is even greater than the sum of the main effects.

8.7: Requirements for Using the *F* Ratio

Objective: **Identify the requirements that need to be considered before the analysis of variance is made**

The analysis of variance should be made only after the researcher has considered the following requirements:

1. *A comparison between three or more independent means.* The *F* ratio is usually employed to make comparisons between three or more means from independent samples. It is possible, moreover, to obtain an *F* ratio rather than a *t* ratio when a two-sample comparison is made. For the two-sample case, $F = t^2$, identical results are obtained. However, a single sample arranged in a panel design (the same group studied at several points in time) cannot be tested in this way. Thus, for example, one may not study improvement in class performance across three examinations during the term using this approach.
2. *Interval data.* To conduct an analysis of variance, we assume that we have achieved the interval level of measurement for the outcome or dependent variable. Categorized or ranked data should not be used as a dependent variable. However, the groups can and are typically formed based on a categorical measure.
3. *Random sampling.* We should have taken our samples at random from a given population of scores.
4. *A normal distribution.* We assume the characteristic we are comparing between and within groups is normally distributed in the underlying populations. Alternatively, the normality assumption will hold if we draw large enough samples from each group.
5. *Equal variances.* The analysis of variance assumes that the population variances for the different groups are all equal. The sample variances, of course, may differ as a result of sampling error. Moderate differences among the sample variances do not invalidate the results of the *F* test. When such differences are extreme (for example, when one of the sample variances is many times larger than another), the *F* test presented here may not be appropriate.

Summary: Analysis of Variance

- Analysis of variance can be used to make comparisons among three or more sample means. Unlike the *t* ratio for comparing only two sample means, the analysis of variance yields an *F* ratio whose numerator represents variation between groups and whose denominator contains an estimate of variation within groups.
- The sum of squares represents the initial step for measuring variation; however, it is greatly affected by sample size. To overcome this problem and control for differences in sample size, we divide the between and within sums of squares by the appropriate degrees of freedom to obtain the mean square.
- The *F* ratio indicates the size of the between-groups mean square relative to the size of the within-groups mean square. The larger the *F* ratio (that is, the larger the between-groups mean square relative to its within-groups counterpart), the more likely we are to reject the null hypothesis and attribute our result to more than just sampling error.

- Earlier, we learned that in studying the difference between two sample means we must compare our calculated *t* ratio against the table *t* (Table B or C). For the purpose of studying differences among three or more means, we now interpret our calculated *F* ratio by comparing it against an appropriate *F* ratio in Table D. On this basis, we decide whether we have a significant difference—whether to retain or reject the null hypothesis.
- After obtaining a significant *F*, we can determine exactly where the significant differences lie by applying Tukey's HSD method for the multiple comparison of means.
- Finally, analysis of variance may be extended to studies of more than one factor. In two-way analysis of variance, we distinguish and test for significant main effects of each factor as well as their interaction. Interaction exists when the effect of two factors combined differs from the sum of their separate (main) effects.

Homework 8.1: Practice Analysis of Variance

1. A researcher is interested in the effect type of residence has on the personal happiness of college students. She selects samples of students who live in campus dorms, in off-campus apartments, and at home and asks the 12 respondents to rate their happiness on a scale of 1 (not happy) to 10 (happy).

Dorms	Apartments	At Home
8	2	5
9	1	4
7	3	3
8	3	4

a. Test the null hypothesis that happiness does not differ by type of residence.

b. Construct a multiple comparison of means by Tukey's method to determine precisely where the significant differences occur.

2. A pediatrician speculated that frequency of visits to her office may be influenced by type of medical insurance coverage. As an exploratory study, she randomly chose 15 patients: 5 whose parents belong to a health maintenance organization (HMO), 5 whose parents had traditional medical insurance, and 5 whose parents were uninsured.

HMO	Traditional	None
12	6	3
6	5	2
8	7	5
7	5	3
6	1	1

a. Using the frequency of visits per year from the above table, test the null hypothesis that type of insurance coverage has no effect on frequency of visits.

b. Conduct a multiple comparison of means by Tukey's method to determine exactly where the significant differences occur.

3. Using the following random sample of death records by neighborhood type, test the null hypothesis that life expectancy in years does not vary by socioeconomic condition of a neighborhood.

Subsidized Housing	Working-Class Neighborhood	Middle-Class Neighborhood
74	82	89
64	71	70
73	76	79
69	80	87
73	79	68

4. A researcher is interested in the effect of employment on satisfaction with marriage. He selects a random sample of 16 married adults who are employed full-time, employed part-time, temporarily unemployed, or chronically unemployed. He asks respondents to rate their satisfaction in marriage on a scale that ranges from 1 (very dissatisfied) to 7 (very satisfied).

Employed Full Time	Employed Part Time	Temporarily Unemployed	Chronically Unemployed
7	4	5	2
5	6	4	0
7	5	4	3
6	4	5	1

a. Using the above data, test the null hypothesis that employment status does not affect marriage satisfaction.

b. Construct a multiple comparison of means by Tukey's HSD method to determine precisely where the significant differences occur.

5. A researcher conducts a study to determine which terrorist organization—Hamas, Hezbollah, or Al Qaeda—poses the greatest threat to innocent civilians. The researcher examines 15 randomly selected bombing attacks committed by these organizations and calculates the number of casualties per attack.

Hamas	Hezbollah	Al Qaeda
6	9	9
2	10	15
3	14	16
7	11	12
8	9	18

a. Using the above casualty data, test the null hypothesis that none of these terrorist groups is more threatening than the rest.

b. Construct a multiple comparison of means by Tukey's HSD method to determine precisely where the significant differences occur.

6. A researcher wishes to determine whether exercising at least three times a week is more effective than pharmaceuticals (antidepressants) or psychotherapy in addressing mental health issues. He randomly assigns 20 psychiatric patients to one of four treatment conditions and asks their self-reported depression at the moment on a scale of 1 (very depressed) to 15 (very happy). Test the null hypothesis that there is no difference between exercise, antidepressants, therapy, and the control group using the following data.

Exercise	Antidepressants	Therapy	Control
9	9	10	7
10	10	11	6
8	14	7	8
6	11	12	6
13	9	8	11

7. Does one's relationship status affect how many vacation days a person takes in a year?

Single	Ongoing Relationship	Civil Union/ Married	Separated	Widowed/ Divorced
11	11	16	23	3
13	33	27	2	21
17	24	18	8	13
9	17	9	8	7
23	29	14	9	2
14	23	24	10	14

a. Using these data collected from samples of adults, test the null hypothesis (that status does not affect the number of vacation days taken).

b. Construct a multiple comparison of means by Tukey's HSD method to determine precisely where the significant differences occur.

8. A health researcher is interested in comparing three methods of weight loss: low-calorie diet, low-fat diet, and low-carb diet. He selects 30 moderately overweight subjects and randomly assigns 10 to each weight-loss program. The following weight reductions (in pounds) were observed after a 1-month period:

Low Calorie		Low Fat		Low Carb	
7	3	7	8	7	14
7	9	8	7	9	10
5	10	8	10	7	11
4	5	9	11	8	5
6	2	5	2	8	6

Test the null hypothesis that the extent of weight reduction does not differ by type of weight-loss program.

9. Consider an experiment to determine the effects of alcohol and marijuana on driving. Five randomly selected subjects are given alcohol to produce legal drunkenness and then are given a simulated driving test (scored from a top score of 10 to a bottom score of 0). Five different randomly selected subjects are given marijuana and then the same driving test. Finally, a control group of five subjects is tested for driving while sober.

Alcohol	Drugs	Control
3	1	8
4	6	7
1	4	8
1	4	5
3	3	6

a. Given the above driving test scores, test for the significance of differences among means of the following groups:

b. Conduct a multiple comparison of means by Tukey's method to determine exactly where the significant differences occur.

10. Using Durkheim's theory of anomie (normlessness) as a basis, a sociologist obtained the following suicide rates (the number of suicides per 100,000 population), rounded to the nearest whole number, for five high-anomie, five moderate-anomie, and five low-anomie metropolitan areas (anomie was indicated by the extent to which newcomers and transients were present in the population):

Anomie		
High	*Moderate*	*Low*
19	15	8
17	20	10
22	11	11
18	13	7
25	14	8

Test the null hypothesis that high-, moderate-, and low-anomie areas do not differ with respect to suicide rates.

11. Psychologists studied the relative efficacy of three different treatment programs—A, B, and C—on illicit drug abuse. The above data represent the number of days of drug abstinence accumulated by 15 patients (5 in each treatment program) for the 3 months after their treatment program ended. Thus, a larger number of days indicates a longer period free of drug use.

Treatment A	Treatment B	Treatment C
90	81	14
74	90	20
90	90	33
86	90	5
75	85	12

a. Test the null hypothesis that these drug-treatment programs do not differ in regard to their efficacy.

b. Conduct a multiple comparison of means by Tukey's method to determine exactly where the significant differences occur.

12. Does a woman's chance of suffering from postpartum depression vary depending on the number of children she already has? To find out, a researcher collected random samples from four groups of women: the first group having just given birth to their first child, the second group having just given birth to their second child, and so on. He then rated their amount of postpartum depression on a scale from 1 to 5 (where 5 = most depression). Test the null hypothesis that the chances of developing postpartum depression do not differ with the number of children to which a woman has previously given birth.

First Child	Second Child	Third Child	Fourth Child
3	3	5	4
2	5	5	3
4	1	3	2
3	3	5	1
2	4	2	5

13. Studies have found that people find symmetrical faces more attractive than faces that are not symmetrical. To test this theory, a psychiatrist selected a random sample of people and showed them pictures of three different faces: a face that is perfectly symmetrical, a face that is slightly asymmetrical, and a face that is highly asymmetrical. She then asked them to rate the three faces in terms of their attractiveness on a scale from 1 to 7, with 7 being the most attractive.

Symmetrical	Slightly Asymmetrical	Highly Asymmetrical
7	5	2
6	4	3
7	5	1
5	2	1
6	4	2
6	5	2

a. Test the null hypothesis that attractiveness does not differ with facial symmetry.

b. Conduct a multiple comparison of means by Tukey's method to determine exactly where the significant differences occur.

14. a. Political theorist Karl Marx is known for his theory that the working class, to put an end to capitalism and establish a communist society, would eventually rise up and overthrow the upper-class members of society who exploit them. One reason for the capitalist workers' discontent, according to Marx's theory, is that these workers take no pride in their work because both the work they do and the products that result belong not to them but to the capitalists they work for. To test this insight, a researcher went to a large factory and interviewed people from three groups—the workers, the managers, and the owners—to see if there is a difference among them in terms of how much pride they take in their work. Given the following scores, with higher scores representing more pride in work, test the null hypothesis that pride in work does not differ by class.

Lower (Workers)	Middle (Managers)	Upper (Owners)
1	4	8
3	7	7
2	5	6
5	6	9
4	8	5
2	6	6
3	5	7

b. Conduct a multiple comparison of means by Tukey's method to determine exactly where the significant differences occur.

15. A psychiatrist wonders if people with panic disorder benefit from one particular type of treatment over any others. She randomly selects patients who have used one of the following treatments: cognitive therapy, behavioral therapy, or medication. She asks them to rate on a scale from 1 to 10 how much the treatment has led to a decrease in symptoms (with a score of 10 being the greatest reduction in symptoms). Test the null hypothesis that the different treatments for panic disorder did not differ in how much they helped these patients.

Cognitive Therapy	Behavior Therapy	Medication
4	6	8
2	3	6
5	4	5
3	8	9
7	6	3
5	4	4
3	7	5

16. Is there a relationship between a mother's education level and how long she breastfeeds her child? A curious researcher selects samples of mothers from three different education levels and determines their length of breastfeeding (measured in months).

Less Than High School	High School Graduate	College Graduate
1.0	1.5	11.0
6.5	4.0	6.5
4.5	3.5	4.5
2.0	1.5	7.5
8.5	5.0	9.0

a. Test the null hypothesis that education level has no effect on how long a mother breastfeeds her child.

b. Conduct a multiple comparison of means by Tukey's method to determine exactly where the significant differences occur.

17. A marriage counselor notices that first marriages seem to last longer than remarriages. To see if this is true, she selects samples of divorced couples from first, second, and third marriages and determines the number of years each couple was married before getting divorced.

First Marriage	Second Marriage	Third Marriage
8.50	7.50	2.75
9.00	4.75	4.00
6.75	3.75	1.50
8.50	6.50	3.75
9.50	5.00	3.50

a. Test the null hypothesis that first, second, and third marriages do not differ by marriage length before divorce.

b. Conduct a multiple comparison of means by Tukey's method to determine exactly where the significant differences occur.

18. In recent years, a number of cases of high school teachers having sexual relationships with their students have made the national news. Interested in how gender combinations influence perceptions of impropriety, a social researcher asks 40 respondents in a survey of education issues about their reaction to a story of a 16-year-old student who is seduced by a 32-year-old teacher. Assigned at random, one-quarter of the respondents are told about a case involving a male teacher and a male student, one-quarter are given a scenario involving a male teacher and a female student, one-quarter are presented a situation of a female teacher with a male student, and one-quarter are presented a female teacher with a female student. All respondents are asked to indicate the level of impropriety from 0 to 10, where 0 is not at all improper and 10 as improper as can be imagined. The results are as follows:

Male Teacher		Female Teacher	
Male Student	*Female Student*	*Male Student*	*Female Student*
10	10	6	5
10	9	7	8
10	9	5	7
9	10	5	9
9	8	2	7
9	6	4	7
10	7	5	5
9	10	7	6
7	9	2	6
9	8	3	10

Using two-way analysis of variance, test gender of teacher, gender of student, and their interaction impact on the level of perceived impropriety surrounding high school teacher–student sexual relationships.

19. How does gender and occupational prestige impact credibility? Graduate students in a public health program are asked to rate the strength of a paper concerning the health risks of childhood obesity. All 30 student raters are given the same paper to evaluate, except that the name and degrees associated with the author have been manipulated by a social researcher. The student raters are randomly assigned to one of six groups, with each group receiving a paper written by a combination of either a male name ("John Forrest") or female name ("Joan Forrest") followed by one of three degrees (M.D., R.N., or Ph.D.). The raters are asked to rate the paper from 1 to 5 on clarity, 1 to 5 on strength of argument, and 1 to 5 on thoroughness. The total rating scores (the sum of the three subscores) are given as follows for each student rater in each of the six groups.

John Forrest			Joan Forrest		
M.D.	R.N.	Ph.D.	M.D.	R.N.	Ph.D.
12	10	10	15	11	11
15	11	8	10	8	11
13	7	13	12	9	12
15	8	12	14	11	8
14	8	9	12	7	8

a. Plot the means for the six groups on a chart with degree on the horizontal axis, mean rating on the vertical axis, and separate lines for each gender.

b. Using two-way analysis of variance, test if gender of author, author's degree, and their interaction impact on the ratings.

20. A clinical researcher wants to determine if antidepressants should be used in conjunction with

therapy to address post-traumatic stress disorder (PTSD) among returning veterans. She designs an experiment with 16 volunteers who had been discharged from the military and diagnosed as suffering from PTSD. After a period of 1 year either in therapy or not and either on antidepressants or not, the researcher measures their PTSD symptoms on a scale ranging from 1 (low) to 15 (high). Using a two-way ANOVA, test the null hypothesis that therapy and antidepressants have no effect, either alone or in combination, on PTSD symptoms.

No Antidepressants		Antidepressants	
No Therapy	Therapy	No Therapy	Therapy
9	9	10	7
10	10	11	6
8	14	7	8
13	9	8	11

Homework 8.2: General Social Survey Practice

1. From the General Social Survey, test whether the number of hours per day that people have to relax (HRSRELAX) varies by marital status (MARITAL). *Hint*: select ANALYZE, COMPARE MEANS, ONE-WAY ANOVA. Remember to weight the cases by WTSSALL. Note also that the Tukey test is selected in the Post Hoc button and that means and standard deviations are selected in the Options button.
2. Based on the General Social Survey, test whether the number of self-reports of poor mental health during the past 30 days (MNTLHLTH) varies by marital status (MARITAL) using one-way ANOVA with the Tukey test and selecting means and standard deviations as an option.
3. Based on the General Social Survey, test whether the number of e-mail hours per week (EMAILHRS) varies by age group (AGE recoded into a new variable AGEGRP) using one-way ANOVA with the Tukey test and selecting means and standard deviations as an option. Before doing your analysis, recode age into a new variable (AGEGRP) using the following groups: 18 to 29; 30 to 44; 45 to 64; 65 or more.
4. Analyze two variables of your choice from the General Social Survey using one-way ANOVA. Remember the dependent variable needs to be an interval/ratio level variable and the independent variable needs to have three or more categories.
5. Reanalyze the example on film content that was presented in this chapter. Notice that the data set is structured with empathy scored on a scale ranging from none (0) to high (10). Sexual content is a categorical variable measured as non-sexual (1) or sexual (2). Violent content is also a categorical variable measure as non-violent (1) or violent (2).

	Film Content	
Empathy	*Sexual*	*Violent*
2	1	1
1	1	1
1	1	1
2	1	1
9	1	2
5	1	2
7	1	2
7	1	2
6	2	1
4	2	1
8	2	1
6	2	1
8	2	2
10	2	2
7	2	2
9	2	2

Use the GLM (General Linear Model) procedure in SPSS, conduct a two-way ANOVA. Hint: ANALYZE, GENERAL LINEAR MODEL, UNIVARIATE. The dependent variable is EMPATHY and the independent variables are entered as fixed factors. The GLM procedure is the easiest method to calculate a two-way ANOVA. It will estimate all main and interaction effects. Compare these SPSS results to the hand calculations from the chapter. Are these results the same within rounding error?

6. Use the GLM procedure in SPSS to test whether respondents' income (CONRINC) varies by marital status (MARITAL) and sex (SEX).
   a. Are the main effects for marital status and sex significant?
   b. Is the interaction effect significant?

Chapter 8 Quiz: Analysis of Variance