CHAPTER 15

TWO- AND THREE-FACTOR ANOVA

The analysis of variance (ANOVA) is not limited to a single factor. The variety and complexity of phenomena and relationships investigated in educational research often dictate the use of two, three, or more factors such as ethnicity, gender, teaching method, demographic area, etc. For example, using one-factor ANOVA to compare ethnic groups on math proficiency does not allow the researcher to test whether the ethnic differences in math proficiency remain the same for males and females. If some (or all) ethnic differences depend on gender, there is an interaction between ethnicity and gender. The presence of such an interaction may mask existing ethnic differences that remain undetected with one-factor ANOVA. In general, an **interaction** between two factors means that the difference in the dependent variable between any two levels of one factor varies across levels of the other factor. Using ANOVA with two or more factors allows the researcher to (a) investigate the role of different sources of error in accounting for the variance in the dependent variable, (b) detect and interpret interactions between factors, and (c) increase the power of tests for differences among factor levels and interactions among factors.

15.1 Two-factor ANOVA

15.1.1 Null Hypotheses in Two-factor ANOVA

Consider again Example 14.1 (Chapter 14, Section 14.4) but this time, in addition to ethnicity, we will also take gender into account to explain differences in math scores. The data layout in Figure 14.2 is now extended to that shown in Figure 15.1 to include two factors: gender and ethnicity. The scores indicate the improvement made by middle schools students on a math test as a result of an experimental school program that incorporates bilingual (English-Spanish) interpretation of mathematics concepts and principles.

In Figure 15.1 (left panel), the column means ($\bar{Y}_{\bullet 1}$, $\bar{Y}_{\bullet 2}$, $\bar{Y}_{\bullet 3}$) are sample estimates of the population means ($\mu_{\bullet 1}$, $\mu_{\bullet 2}$, $\mu_{\bullet 3}$) for the three ethnic groups, whereas the row means ($\bar{Y}_{1\bullet}$, $\bar{Y}_{2\bullet}$) are sample estimates of the population means ($\mu_{1\bullet}$, $\mu_{2\bullet}$) for females and males, respectively. The sample grand mean ($\bar{Y}_{\bullet\bullet}$) is an estimate of the population grand mean ($\mu_{\bullet\bullet}$). The SPSS entry (right panel) shows the scores and the coding values for Ethnicity (1 = Caucasian, 2 = African-American, 3 = Hispanic) and Gender (0 = female, 1 = male). Under this **two-factor ANOVA** (or **two-way ANOVA**) model, the following three null hypotheses are testable:

H_{01}: $\mu_{1\bullet} = \mu_{2\bullet}$; [**No gender differences** (ethnicity ignored)]
H_{02}: $\mu_{\bullet 1} = \mu_{\bullet 2} = \mu_{\bullet 3}$; [**No ethnic differences** (gender ignored)]
H_{03}: **There is NO interaction between *Gender* and *Ethnicity***.

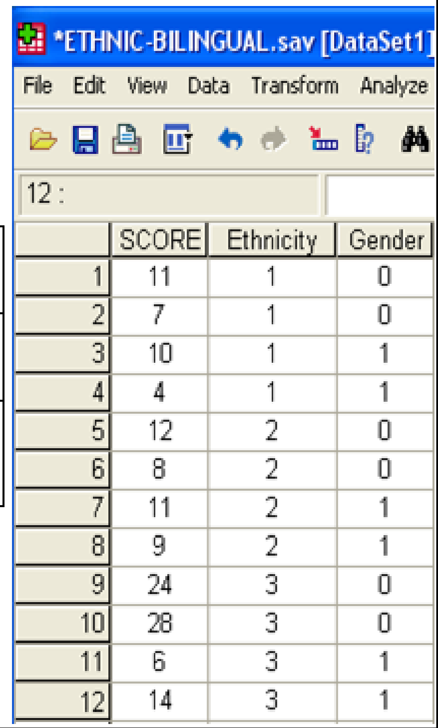
In the general case of two factors, with *J* levels in Factor A and *K* levels in Factor B, the null hypotheses with the (A x B) two-factor ANOVA are:

H_{01}: $\mu_{1\bullet} = \mu_{2\bullet} = \ldots = \mu_{J\bullet}$ **(15.1)**
H_{02}: $\mu_{\bullet 1} = \mu_{\bullet 2} = \ldots = \mu_{\bullet K}$ **(15.2)**
H_{03}: There is NO interaction between factors A and B. **(15.3)**

When H_{01} is rejected, we say that there is a statistically significant **main effect** of factor A. Likewise, when H_{02} is rejected, there is a statistically significant **main effect** of factor B. When H_{03} is rejected, there is a statistically significant **interaction** between the factors A and B. The notation (J x K) ANOVA indicates that this is a two-factor ANOVA with J levels in the first factor and K levels in the second factor. For example, Figure 15.1 depicts a 2 x 3 ANOVA (J = 2, K = 3), which can be referred to also as *Gender* x *Ethnicity* ANOVA. The order of the factors is irrelevant, but typically A x B means that A is the "row" factor and B is the "column" factor.

Figure 15.1 *Data layout for two-factor (gender x ethnicity) ANOVA*

	ETHNICITY			
GENDER	Caucasian	Afr-Amer	Hispanic	
Female	11, 7	12, 8	24, 28	$\bar{Y}_{1\bullet} = 15$
Male	10, 4	11, 9	6, 14	$\bar{Y}_{2\bullet} = 9$
	$\bar{Y}_{\bullet 1} = 8$	$\bar{Y}_{\bullet 2} = 10$	$\bar{Y}_{\bullet 3} = 18$	$\bar{Y}_{\bullet\bullet} = 12$

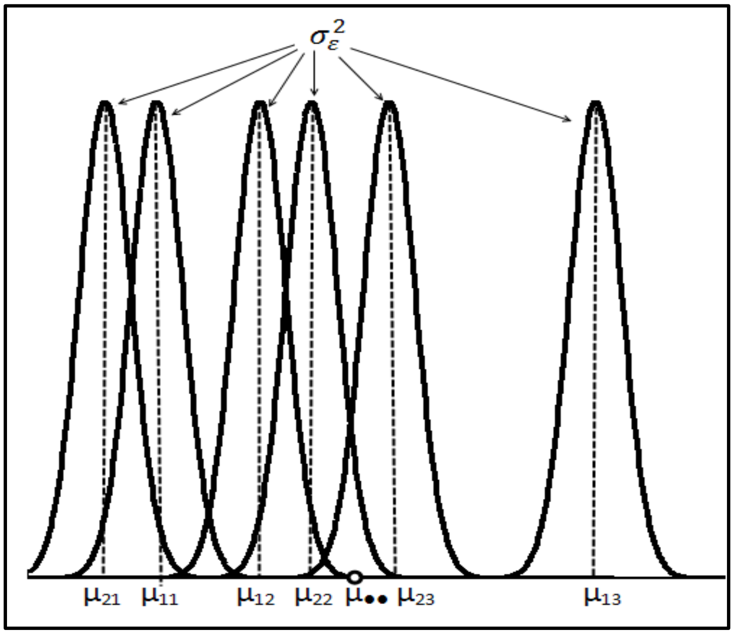


	SCORE	Ethnicity	Gender
1	11	1	0
2	7	1	0
3	10	1	1
4	4	1	1
5	12	2	0
6	8	2	0
7	11	2	1
8	9	2	1
9	24	3	0
10	28	3	0
11	6	3	1
12	14	3	1

15.1.2 Assumptions in Two-factor ANOVA

The assumptions of *normality*, *homogeneity of variance*, and *independence* in one-factor ANOVA remain in two-factor ANOVA. The difference is that the population distribution and its error variance, σ_ε^2, are not the same in one- and two-factor ANOVA models. For the ANOVA model in Figure 15.1, for example, there are six population distributions of scores for students with the same gender and ethnicity. That is, the two scores in each cell represent a random sample of two observations (n = 2) from the population distribution of scores for students with the same gender and ethnicity. For example, the cell mean for Caucasian females (first row, first column), $\bar{Y}_{11}$= 9, is a sample estimate of the population mean for all Caucasian females, μ_{11}. Likewise, the other five cell means, $\bar{Y}_{12}$= 10, $\bar{Y}_{13}$= 26, $\bar{Y}_{21}$= 7, $\bar{Y}_{22}$= 10, and $\bar{Y}_{23}$= 10 are estimates of the population cell means, μ_{12}, μ_{13}, μ_{21}, μ_{22}, and μ_{23}, respectively. Under the ANOVA assumptions of normality and homogeneity of variance, all six population distributions are normal and have equal variances, $\sigma_{11}^2 = \sigma_{12}^2 = \sigma_{13}^2 = \sigma_{21}^2 = \sigma_{22}^2 = \sigma_{23}^2$, denoted σ_ε^2 (*population error variance*) — see Figure 15.2 [compare to Figure 14.1].

Figure 15.2 *Population distributions for 2 x 3 ANOVA*



15.1.3 Effects in Two-factor ANOVA

The concept of effects in one-factor ANOVA, introduced in Chapter 14 (Section 14.3), is extended here to that of effects in two-factor ANOVA. Consider a two-factor (A x B) ANOVA with *J* levels for factor A and *K* levels for factor B. If depicted in a two-way (*J* x *K*) ANOVA table, there are *J* horizontal rows (factor A), *K* columns (factor B), and *JK* cells. There are three types of effects with the two-factor ANOVA:

- **Row effects** (for the levels of factor A):

$$\alpha_j = \mu_{j\bullet} - \mu_{\bullet\bullet}\ ;\ (j = 1, 2, \ldots, J) \quad \textbf{(15.4)}$$

- **Column effects** (for the levels of factor B)

$$\beta_k = \mu_{\bullet k} - \mu_{\bullet\bullet}\ ;\ (k = 1, 2, \ldots, K) \quad \textbf{(15.5)}$$

- **Cell effects** (interaction terms):

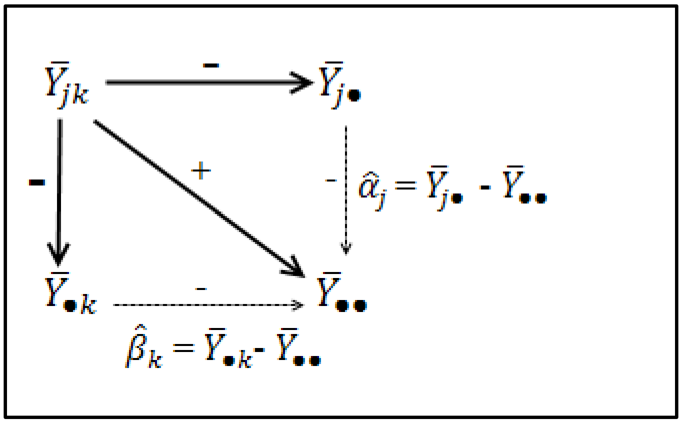
$$\alpha\beta_{jk} = \mu_{jk} - (\mu_{\bullet\bullet} + \alpha_j + \beta_k) \quad \textbf{(15.6)}$$

After replacing the values of $\alpha_{j\bullet}$ and $\beta_{\bullet k}$ from (15.4) and (15.5) into (15.6) and using some simple algebra, the interaction term can be represented as follows:

$$\alpha\beta_{jk} = \mu_{jk} - \mu_{j\bullet} - \mu_{\bullet k} + \mu_{\bullet\bullet} \quad \textbf{(15.7)}$$

As Equation 15.6 shows, the interaction term $\alpha\beta_{jk}$ is obtained by subtracting from the cell mean (μ_{jk}) the total effect ($\mu_{\bullet\bullet}$), the row effect (α_j), and the column effect (β_k). Thus, the interaction term is obtained from the cell mean after "partialling" out all other effects (total effect, row effect, and column effect) that affect the cell mean. While Equation 15.6 clarifies the meaning of interaction terms, $\alpha\beta_{jk}$, Equation 15.7 is more convenient for computations of their sample estimate (see Figure 15.3).

Figure 15.3 *Estimation of effects in two-factor ANOVA*



Thus, α_j is the effect produced by the row mean $\mu_{j\bullet}$, β_k is the effect produced by the column mean $\mu_{\bullet k}$, and $\alpha\beta_{jk}$ is the effect produced by the cell mean μ_{jk}. If Y_{ijk} is the score of person *i* in the cell located at row *j* and column *k*, the deviation of this score from the cell mean is the *error term* for the score, $\varepsilon_{ijk} = Y_{ijk} - \mu_{jk}$, referred to also as the **random effect** of score Y_{ijk}.

In the general case of *J* levels in Factor A and *K* levels in Factor B, each "effect" represents a deviation from the mean and, therefore, the sum of all row effects equals zero ($\sum \alpha_j = 0$), the sum of all column effects equals zero ($\sum \beta_k = 0$), and the sum of all interaction terms equals zero ($\sum\sum \alpha\beta_{jk} = 0$); ($j = 1, 2, \ldots, J$; $k = 1, 2, \ldots, K$). However, the sum of squared effects for the levels of a factor (or interaction between factors) equals zero *only if* the null hypothesis for this factor (or interaction between factors) is true. Thus, the null hypotheses with the (A x B) two-factor ANOVA (see Equations 15.1-15.3) can be represented by using row effects, column effects, and interaction terms, respectively:

Main effect of factor A, H_{01}: $\sum \alpha_j^2 = 0$, **(15.8)**

Main effect of factor B, H_{02}: $\sum \beta_k^2 = 0$, **(15.9)**

Interaction between factors A and B, H_{03}: $\sum\sum \left(\alpha\beta_{jk}\right)^2 = 0$, **(15.10)**

where the summations are for $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$.

EXAMPLE 15.1 This example illustrates how to calculate sample estimates of row effects, column effects, and interaction terms in two-factor ANOVA using the data in Figure 15.1. The estimation of error terms is also illustrated. Using the computation formulas provided in Figure 15.3, we obtain sample estimates of the effects of gender levels (row effects), ethnic levels (column effects), and interaction terms as follows:

Effects of gender levels (row effects): $\hat{\alpha}_1 = \bar{Y}_{1\bullet} - \bar{Y}_{\bullet\bullet} = 15 - 12 = 3$,

$\hat{\alpha}_2 = \bar{Y}_{2\bullet} - \bar{Y}_{\bullet\bullet} = 9 - 12 = -3$.

Effects of ethnic levels (column effects): $\hat{\beta}_1 = \bar{Y}_{\bullet 1} - \bar{Y}_{\bullet\bullet} = 8 - 12 = -4$,

$\hat{\beta}_2 = \bar{Y}_{\bullet 2} - \bar{Y}_{\bullet\bullet} = 10 - 12 = -2$,

$\hat{\beta}_3 = \bar{Y}_{\bullet 3} - \bar{Y}_{\bullet\bullet} = 18 - 12 = 6$.

Interaction terms: $\widehat{\alpha\beta}_{11} = \bar{Y}_{11} - \bar{Y}_{1\bullet} - \bar{Y}_{\bullet 1} + \bar{Y}_{\bullet\bullet} = 9 - 15 - 8 + 12 = 21 - 23 = -2$,

$\widehat{\alpha\beta}_{12} = \bar{Y}_{12} - \bar{Y}_{1\bullet} - \bar{Y}_{\bullet 2} + \bar{Y}_{\bullet\bullet} = 10 - 15 - 10 + 12 = 22 - 25 = -3$,

$\widehat{\alpha\beta}_{13} = \bar{Y}_{13} - \bar{Y}_{1\bullet} - \bar{Y}_{\bullet 3} + \bar{Y}_{\bullet\bullet} = 26 - 15 - 18 + 12 = 38 - 33 = 5$,

$\widehat{\alpha\beta}_{21} = \bar{Y}_{21} - \bar{Y}_{2\bullet} - \bar{Y}_{\bullet 1} + \bar{Y}_{\bullet\bullet} = 7 - 9 - 8 + 12 = 19 - 17 = 2$,

$\widehat{\alpha\beta}_{22} = \bar{Y}_{22} - \bar{Y}_{2\bullet} - \bar{Y}_{\bullet 2} + \bar{Y}_{\bullet\bullet} = 10 - 9 - 10 + 12 = 22 - 19 = 3$,

$\widehat{\alpha\beta}_{23} = \bar{Y}_{23} - \bar{Y}_{2\bullet} - \bar{Y}_{\bullet 3} + \bar{Y}_{\bullet\bullet} = 10 - 9 - 18 + 12 = 22 - 27 = -5$.

The estimation of error terms for an individual observation, $\varepsilon_{ijk} = Y_{ijk} - \mu_{jk}$, is performed by replacing the population cell mean, μ_{jk}, with its sample estimate, $\bar{Y}_{jk}$. Thus, $\hat{\varepsilon}_{ijk} = Y_{ijk} - \bar{Y}_{jk}$. For example, the error terms for the two observations in the first cell (i = 1, k = 1) in Figure 15.1 are: $\hat{\varepsilon}_{111} = Y_{111} - \bar{Y}_{11} = 11 - 9 = 2$ and $\hat{\varepsilon}_{211} = Y_{211} - \bar{Y}_{11} = 7 - 9 = -2$. Likewise, the error terms of the observations for the "Male-Hispanic" cell (i = 2, k = 3) are: $\hat{\varepsilon}_{123} = Y_{123} - \bar{Y}_{23} = 6 - 10 = -4$ and $\hat{\varepsilon}_{223} = Y_{223} - \bar{Y}_{23} = 14 - 10 = 4$, etc. Note that the sum of error terms within a cell equals zero, and that the sum of all row effects equals zero ($\sum \hat{\alpha}_j = 3 - 3 = 0$), the sum of all column effects equals zero ($\sum \hat{\beta}_k = -4 - 2 + 6 = 0$), and the sum of all interaction terms equals zero ($\sum\sum \widehat{\alpha\beta}_{jk} = -2 - 3 + 5 + 2 + 3 - 5 = 0$).

15.1.4 Linear Model for the Data in Two-factor ANOVA

The **linear model** for the data in two-factor ANOVA is an extension of its counterpart for one-factor ANOVA (see Chapter 14, Section 14.5) to accommodate for the presence of two factors and a possible interaction between them:

$$Y_{ijk} = \mu_{\bullet\bullet} + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk}, \qquad \textbf{(15.11)}$$

where Y_{ijk} is the score of person i in the cell located at row j and column k, and ε_{ijk} is the error term of Y_{ijk} ($\varepsilon_{ijk} = Y_{ijk} - \mu_{jk}$). The variance of error terms, ε_{ijk}, across all observations for the population within a cell is the *population error variance*, σ_ε^2 (see Figure 15.2). According to the ANOVA linear model in Equation 15.11, each individual score can be represented as a sum of five effects: total effect ($\mu_{\bullet\bullet}$), row effect ($\alpha_{j\bullet}$), column effect ($\beta_{k\bullet}$), interaction effect ($\alpha\beta_{jk}$), and random effect (ε_{ijk}). By moving the grand mean ($\mu_{\bullet\bullet}$) from the right-hand side in Equation 15.11 into its left-hand side (with an opposite sign), we obtain the linear model for the deviation of individual scores from the grand mean:

$$Y_{ijk} - \mu_{\bullet\bullet} = \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk}. \qquad \textbf{(15.12)}$$

The deviation ($Y_{ijk} - \mu_{\bullet\bullet}$) is the building unit of the total variability in the dependent variable, Y. By squaring both sides of Equation 15.12 and using some summation algebra, we can see that the total variation in Y equals the sum of the squared effects of rows, columns, interaction terms, and error terms:

$$\sum\sum\sum\left(Y_{ijk} - \mu_{\bullet\bullet}\right)^2 = \sum \alpha_j^2 + \sum \beta_k^2 + \sum\sum\left(\alpha\beta_{jk}\right)^2 + \sum\sum\sum \varepsilon_{ijk}^2 \qquad \textbf{(15.13)}$$

where the triple summation is across subjects (i), rows (j), and columns (k). Note that the first three terms in the right-hand side in Equation 15.13 are used to represent the three null hypotheses in two-factor ANOVA (Equations 15.8, 15.9, and 15.10, respectively).

15.1.5 Sum of Squares in Two-factor ANOVA

Let's assume that the two-factor (A x B) ANOVA design is *balanced*, with *n* observations in each cell (*j*, *k*), where $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$. This simplifies the analytic presentation without any loss of generality for the case when the ANOVA design is unbalanced. The
sample estimate of the sum in the left-hand side in Equation 15.13 is the **total sum of squares** (SS_T). That is,

$$SS_T = \sum\sum\sum\left(Y_{ijk} - \bar{Y}_{\bullet\bullet}\right)^2, \qquad \textbf{(15.14)}$$

where the triple summation is across $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, J$, and $k = 1, 2, \ldots, K$. The four components of the SS_T (estimates of the four terms in the right-hand side in Equation 15.13) are:

- **Sum of squares for factor A**:

$$SS_A = (nK)\sum \hat{\alpha}_j^2 = (nK)\sum\left(\bar{Y}_{j\bullet} - \bar{Y}_{\bullet\bullet}\right)^2, \qquad \textbf{(15.15)}$$

where (nK) is the number of observations used in the computation of the row mean, $\bar{Y}_{j\bullet}$.

- **Sum of squares for factor B**:

$$SS_B = (nJ)\sum \hat{\beta}_k^2 = (nJ)\sum(\bar{Y}_{\bullet k} - \bar{Y}_{\bullet\bullet})^2, \qquad \textbf{(15.16)}$$

where (nJ) is the number of observations used in the computation of the column mean, $\bar{Y}_{\bullet k}$.

- **Sum of squares for the A x B interaction:**

$$SS_{AB} = n\sum\sum\left(\widehat{\alpha\beta}_{jk}\right)^2 = n\sum\sum\left(\bar{Y}_{jk} - \bar{Y}_{j\bullet} - \bar{Y}_{\bullet k} + \bar{Y}_{\bullet\bullet}\right)^2, \qquad \textbf{(15.17)}$$

where n is the number of observations used in the computation of the cell mean, $\bar{Y}_{jk}$.

- **Sum of squares within cells:**

$$SS_W = \sum\sum\sum \hat{\varepsilon}_{ijk}^2 = \sum\sum\sum\left(Y_{ijk} - \bar{Y}_{jk}\right)^2 \qquad \textbf{(15.18)}$$

Thus, the sample-based equivalent of Equation 15.13 for partitioning the total variability into four components is

$$\mathbf{SS_T = SS_A + SS_B + SS_{AB} + SS_W}. \qquad \textbf{(15.19)}$$

EXAMPLE 15.2 This example illustrates how to calculate the sum of squares for the two-factor ANOVA data in Figure 15.1, where factor A = *Gender* and Factor B = *Ethnicity*. Using Equations 15.15, 15.6, and 15.7, with some intermediate results from Example 15.1, two observations per cell ($n = 2$), two levels of factor A ($J = 2$), and three levels of factor B ($K = 3$), we obtain

$$SS_A = (nK)\sum \hat{\alpha}_j^2 = (2)(3)(\hat{\alpha}_1^2 + \hat{\alpha}_2^2) = (6)[3^2 + (-3)^2] = (6)(9 + 9) = (6)(18) = 108,$$

$SS_B = (nJ)\sum \beta_k^2 = (2)(2)(\beta_1^2 + \beta_2^2 + \beta_3^2) = (4)[(-4)^2 + (-2)^2 + 6^2] = (4)(16 + 4 + 36) = 224,$

$SS_{AB} = n\sum\sum(\widehat{\alpha\beta}_{jk})^2 = (2)[(\widehat{\alpha\beta}_{11})^2 + (\widehat{\alpha\beta}_{12})^2 + (\widehat{\alpha\beta}_{13})^2 + (\widehat{\alpha\beta}_{21})^2 + (\widehat{\alpha\beta}_{22})^2 + (\widehat{\alpha\beta}_{23})^2]$

$= (2)[(-2)^2 + (-3)^2 + 5^2 + 2^2 + 3^2 + (-5)^2] = (2)(4 + 9 + 25 + 4 + 9 + 25) = (2)(76) = 152.$

To calculate $SS_W = \sum\sum\sum \hat{\varepsilon}_{ijk}^2$, we use the values provided in Example 15.1 of four error terms in two cells [$\hat{\varepsilon}_{111} = 2$, $\hat{\varepsilon}_{211} = -2$, $\hat{\varepsilon}_{123} = -4$, $\hat{\varepsilon}_{223} = 4$] and the values of the eight error terms in the remaining four cells [the computation of which is not shown here for space consideration] thus obtaining the sum of squares within cells:

$SS_W = \hat{\varepsilon}_{111}^2 + \hat{\varepsilon}_{211}^2 + \ldots + \hat{\varepsilon}_{123}^2 + \hat{\varepsilon}_{223}^2$

$= 2^2 + (-2)^2 + 2^2 + (-2)^2 + (-2)^2 + 2^2 + 3^2 + (-3)^2 + 1^2 + (-1)^2 + (-4)^2 + 4^2 = 76.$

As all terms in the right-hand side of Equation15.19 are now known, we can compute the *sum of squares total*:

$SS_T = SS_A + SS_B + SS_{AB} + SS_W = 108 + 224 + 152 + 76 = 560.$

Thus, we have: $SS_A = 108$, $SS_B = 224$, $SS_{AB} = 152$, $SS_W = 76$, and $SS_T = 560$.

15.1.6 Mean Squares in Two-factor ANOVA

By taking the ratio of each sum of squares (SS) to its degrees of freedom (*df*), we can obtain the **mean squares** (**MS**) **—** estimates of the population variances of the effects for factor A, factor B, their interaction (A x B), and the population variance, σ_ε^2. The degrees of freedom for the sum of squares (SS_A, SS_B, SS_{AB}, and SS_W) are obtained as follows:

$$df_A = J - 1,\ df_B = K - 1,\ df_{AB} = (J - 1)(K - 1),\text{ and } df_W = JK(n - 1) \qquad \textbf{(15.20)}$$

Thus, the corresponding mean squares are:

$$MS_A = \frac{SS_A}{J-1},\ MS_B = \frac{SS_B}{K-1},\ MS_{AB} = \frac{SS_{AB}}{(J-1)(K-1)},\text{ and } MS_W = \frac{SS_W}{JK(n-1)}. \qquad \textbf{(15.21)}$$

EXAMPLE 15.3 This example illustrates how to calculate the mean squares for the two-factor ANOVA data in Figure 15.1. Using the formulas for mean squares in (15.21), with $n = 2$, $J = 2$, $K = 3$, and the sum of squares obtained in Example 15.2 ($SS_A = 108$, $SS_B = 224$, $SS_{AB} = 152$, and $SS_W = 76$), we have:

$$MS_A = \frac{SS_A}{J-1} = \frac{108}{2-1} = 108,$$

$$MS_B = \frac{SS_B}{K-1} = \frac{224}{3-1} = 112,$$

$$MS_{AB} = \frac{SS_{AB}}{(J-1)(K-1)} = \frac{152}{(2-1)(3-1)} = 76,\text{ and}$$

$$MS_W = \frac{SS_W}{JK(n-1)} = \frac{76}{(2)(3)(2-1)} = 12.667.$$

The within-cells variance for the sample, MS_W, is entirely random, thus representing an estimate of the population error variance, σ_ε^2. In general, MS_A consists of two parts (variance components): a random part and a part that is due to differences (if any) among the levels of factors A. Likewise, MS_B consists of two parts: a random part and a part due to differences (if any) among the levels of factors B. Finally, MS_{AB} also consists of a random part and a part due to interaction (if any) between the two factors, A and B.

15.1.7 Testing the Null Hypotheses in Two-factor ANOVA

As discussed earlier, three null hypotheses are testable in two-factor ANOVA — two main effects and an interaction effect (see, 15.8, 15.9, and 15.10):

Main effect of factor A, H_{01}: $\sum \alpha_j^2 = 0$,
Main effect of factor B, H_{02}: $\sum \beta_k^2 = 0$,
Interaction between factors A and B, H_{03}: $\sum\sum(\alpha\beta_{jk})^2 = 0$.

The logic behind the F-test for the null hypothesis in one-factor ANOVA (Chapter 14, Section 14.6) carries over into testing the three null hypotheses in two-factor ANOVA. For example, if H_{01} is true, MS_A does not contain a variance component due to differences among the levels in factor A and, therefore, consists only of a random variance component. In other words, when H_{01} is true, MS_A represents a sample estimate of the population error variance, σ_ε^2. Thus, when H_{01} is true, MS_A and MS_W represent two sample estimates of the same population variance, σ_ε^2, and, therefore, their ratio must follow the F-distribution. This yields the following F-statistic for H_{01} (main effect of factor A):

$$F_A = \frac{MS_A}{MS_W} \tag{15.22}$$

Likewise, the following F-statistics are used to test H_{02} (the main effect of factor A) and H_{03} (the interaction between factors A and B), respectively:

$$F_B = \frac{MS_B}{MS_W} \text{ and } F_{AB} = \frac{MS_{AB}}{MS_W} \tag{15.23}$$

> **NOTE [15.1] The *F*-test for a specific main effect in two-factor ANOVA is generally more powerful than the *F*-test for that effect in one-factor ANOVA.** This is because the MS_W with the former (*within-cell variance*) is smaller than the MS_W with the latter (*within-column variance*).

EXAMPLE 15.4 This example illustrates how to calculate the F-ratios and how to use them in testing the null hypotheses for the two-factor ANOVA data in Figure 15.1 (factor A = *Gender* and Factor B = *Ethnicity*) at the .05 level of significance. The values of the dependent variable in this case are scores that indicate the improvement made by middle school students on a math test as a result of an experimental school program that incorporates bilingual (English-Spanish) interpretation of mathematics concepts and principles (see Chapter 14, Example 14.1). We will briefly refer to these scores here as "math gain scores."

As the mean squares for these data are already known from Example 15.3 (MS_A = 108, MS_B = 112, MS_{AB} = 76, and MS_W = 12.667), we use directly Formulas 15.22 and 15 as follows:

$$F_A = \frac{MS_A}{MS_W} = \frac{108}{12.667} = 8.526;\ F_B = \frac{MS_B}{MS_W} = \frac{112}{12.667} = 8.842;\text{ and } F_{AB} = \frac{MS_{AB}}{MS_W} = \frac{76}{12.667} = 6.00$$

Using Formulas 15.20, we compute the degrees of freedom (*df*) for each F-ratio. Note that the two-factor ANOVA design in Figure 15.1 is balanced, with two observations per cell ($n = 2$), two levels of factor A ($J = 2$), and three levels of factor B ($K = 3$). First, all F-ratios have the same denominator (MS_W) with 6 degrees of freedom ($df_W = 6$). That is, $df_W = JK(n - 1) = (2)(3)(2 - 1) = 6$. The degrees of freedom for the numerators of F-ratios are: $df_A = J - 1 = 2 - 1 = 1$, $df_B = K - 1 = 3 - 1 = 2$, and $df_{AB} = (J - 1)(K - 1) = (2 - 1)(3 - 1) = 2$.

The ANOVA null hypotheses (H_{01}, H_{02}, and H_{03}) in this example are:

Main effect of gender, H_{01}: $\mu_{1\bullet} = \mu_{2\bullet}$ [or, H_{01}: $\sum \alpha_j^2 = 0$; ($j = 1, 2$)]
Main effect of ethnicity, H_{02}: $\mu_{\bullet 1} = \mu_{\bullet 2} = \mu_{\bullet 3}$ [or, H_{02}: $\sum \beta_k^2 = 0$; ($k = 1, 2, 3$)]
Interaction (*gender* x *ethnicity*), H_{03}: $\sum\sum\left(\alpha\beta_{jk}\right)^2 = 0$; ($j = 1, 2$; $k = 1, 2, 3$).

The first null hypothesis, H_{01}, is tested by comparing the F_A test value (8.526) against the critical F-value for $\alpha = .05$, $df_A = 1$, and $df_W = 6$ [see Table A-4]. Because the computed F_A value (8.526) exceeds the critical F-value (5.99), we reject H_{01}. Thus, there is a statistically significant main effect of gender at the .05 level. In other words, males and females differ on their mean scores on the dependent variable (math gain score).

The second null hypothesis, H_{02}, is also rejected since the computed F-value ($F_B = 8.842$) exceeds the critical F-value (5.14) for $\alpha = .05$, $df_B = 2$, and $df_W = 6$. Thus, there is a statistically significant difference between the means of at least two of the three ethnic groups. Which particular groups are different can be determined using multiple comparisons (e.g., the Tukey post hoc test). This is illustrated in Example 15.5.

Finally, H_{03} is also rejected since the computed F-value ($F_{AB} = 6.00$) is greater than the critical F-value (5.14) for $\alpha = .05$, $df_{AB} = 2$, and $df_W = 6$. This indicates that there is a statistically significant interaction between gender and ethnicity.

Note that the two-factor ANOVA test in this example provides evidence of differences among the three ethnic groups, whereas this was not the case with the one-factor ANOVA for the same data in Example 14.2 (Chapter 14). This occurred because the computed F-value here is $F_B = MS_B/MS_W = 108/12.667 = 8.842$, whereas the F-value with the one-factor ANOVA is smaller: $F_B = MS_B/MS_W = 108/37.333 = 3.00$. The larger F-ratio with the two-factor ANOVA (8.842) is due to its smaller denominator ($MS_W = 12.667$) compared to that with the one-factor ANOVA reported in Figure 14.3 ($MS_W = 37.333$) — see NOTE [15.1].

15.1.8 Omnibus Effect Size in Two-factor ANOVA

The measure of omnibus effect size, η^2 (eta squared) presented in Chapter 14 (Section 14.8.2) is applicable for the main and interaction effects in a two-factor (A x B) ANOVA.

$$\eta^2 = \frac{SS_{effect}}{SS_T}, \tag{15.24}$$

where SS_T is the *sum of squares total* (see Equation 15.14) and SS_{effect} is the sum of squares of the "effect" A, B, or A x B. That is, (a) for factor A: $SS_{effect} = SS_A$ (Equation 15.15), (b) for factor B: $SS_{effect} = SS_B$ (Equation 15.16), and (c) for the (A x B) interaction: $SS_{effect} = SS_{AB}$

(Equation 15.17). For example, for the two factor (*Gender* x *Ethnicity*) ANOVA in Examples 15.2, 15.3, and 15.4, we have $SS_A = 108$, $SS_B = 224$, $SS_{AB} = 152$, and $SS_W = 76$. Using Equation 15.19, we obtain: $SS_T = SS_A + SS_B + SS_{AB} + SS_W = 108 + 224 + 152 + 76 = 560$. Given this, we use Equation 15.24 to obtain the effect size for factor A (η^2_A), factor B (η^2_B), and the interaction between them (η^2_{AB}) as follows:

$$\eta^2_A = \frac{SS_A}{SS_T} = \frac{108}{560} = .193,\ \eta^2_B = \frac{SS_B}{SS_T} = \frac{224}{560} = .400, \text{ and } \eta^2_{AB} = \frac{SS_{AB}}{SS_T} = \frac{152}{560} = .271.$$

Thus, following the Cohen's guidelines described in NOTE [14.4], we find that all three effect sizes are large.

In ANOVA with two or more factors, an effect size measure called **partial eta squared** ($p\eta^2$) is also used to indicate the proportion of the variability in the dependent variable, *Y*, accounted for by a given effect (A, B, or A x B) after "partialling out" (controlling for) the contribution of all other effects. The formula for partial eta squared is

$$p\eta^2 = \frac{SS_{effect}}{SS_{\text{effect}} + SS_W}, \tag{15.25}$$

where SS_{effect} is the same as with η^2 (i.e., SS_A, SS_B, or SS_{AB}) and SS_W is the *sum of squares within cells* (or *sum of squares error*) — see Equation 15.18.

Using the sum of squares values with the above calculations of the effects η^2_A, η^2_B, and η^2_{AB}, we use Formula 15.25 to compute their partial counterparts as follows:

$$p\eta^2_A = \frac{SS_A}{SS_A + SS_W} = \frac{108}{108+76} = .587,$$

$$p\eta^2_B = \frac{SS_B}{SS_B + SS_W} = \frac{224}{224+76} = .747, \text{ and}$$

$$p\eta^2_{AB} = \frac{SS_{AB}}{SS_{AB} + SS_W} = \frac{152}{152+76} = .667.$$

NOTE [15.2] Eta squared for an effect is always smaller than its partial eta squared ($\eta^2 < p\eta^2$). This is because the denominator for η^2 (SS_T) in Formula 15.24 is larger than the denominator for $p\eta^2$ ($SS_{effect} + SS_W$) in Formula 15.25, as the latter is a part of the former (see Equation 15.19). Also, **the Cohen's guidelines for (small, medium, and large) effect size do not apply for the partial eta squared** because its denominator changes across effects (e.g., A, B, and A x B, as shown here above).

In the context of two-factor ANOVA, the "translation" of Formula 14.34 (in Chapter 14) for **omega squared** (ω^2) — the effect size adjusted for the population — is straightforward:

$$\omega^2 = \frac{SS_{effect} - (df_{effect})MS_W}{SS_T + MS_W}, \tag{15.26}$$

where df_{effect} is the degrees of freedom for the effect (see Formulas 15.20).

For example, given that $SS_A = 108$, $SS_B = 224$, and $SS_{AB} = 152$ [which we already used in the computations of η^2 and $p\eta^2$], we can now use Formula 15.26 to compute omega squared for factor A (ω_A^2), factor B (ω_A^2), and their interaction (ω_{AB}^2) as follows:

$$\omega_A^2 = \frac{SS_A - (df_A)MS_W}{SS_T + MS_W} = \frac{108 - (1)(76)}{560 + 76} = .050,$$

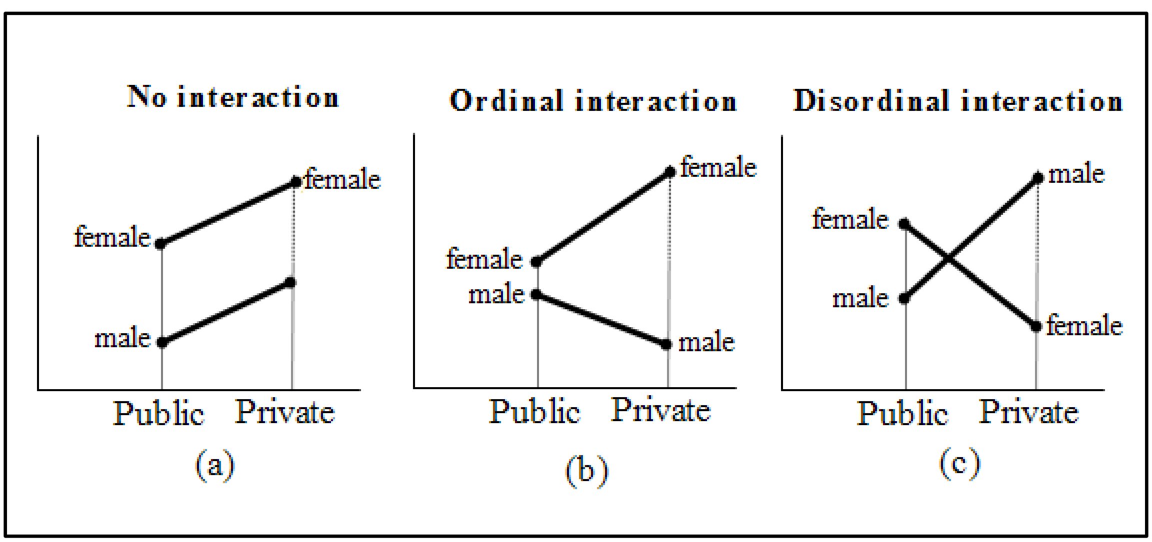
$$\omega_B^2 = \frac{SS_B - (df_B)MS_W}{SS_T + MS_W} = \frac{224 - (2)(76)}{560 + 76} = .113, \text{ and}$$

$$\omega_{AB}^2 = \frac{SS_{AB} - (df_{AB})MS_W}{SS_T + MS_W} = \frac{152 - (2)(76)}{560 + 76} = .00.$$

15.1.9 Types of Interaction in Two-factor ANOVA

There are two types of interaction that may occur in two-factor ANOVA — ordinal or disordinal interaction. **Ordinal interaction** is when the order of the mean scores for the levels in one factor is the same across the levels of other factor. **Disordinal interaction** is when the order of mean scores for the levels in one factor is not the same across the levels of other factor. The three plots in Figure 15.4 illustrate the hypothetical cases of (a) no interaction, (b) an ordinal interaction, and (c) a disordinal interaction between two factors — *gender* (male, female) and type of school (public, private), where the dependent variable, Y, is attitude toward school. The cell means, $\bar{Y}_{jk}$, are presented on the vertical axis. Keep in mind, however, that in a real study there must be evidence that the F-test for interaction is statistically significant in order to proceed with the interpretation of the interaction plot (see F_{AB} in Equation 15.23).

Figure 15.4 *Three hypothetical cases of (a) no interaction, (b) ordinal interaction, and (c) disordinal interaction) in a two-factor ANOVA with factors gender (male, female) and type of school (public, private)*



15.1.10 Testing for Simple Main Effects

When the interaction in a two-factor ANOVA is statistically significant, the interpretation of main effects is of little interest regardless of their statistical significance. For example, with the interactions depicted in Figure 15.4 (b and c) there would likely be no statistically significant main effect of gender. Clearly, interpreting the (lack of) main effect for gender in each of these two scenarios would mask existing gender differences across public and private schools. It would be important, therefore, to test for gender differences separately for public and private schools. This is referred to as testing for *simple main effects* of gender across public and private schools.

In general, if the (A x B) interaction in a two-factor ANOVA is statistically significant, **testing for simple main effects** of factor A is performed by testing the difference between any two levels of factor A across the levels of factor B. Likewise, one can perform testing for simple main effects of factor B across the levels of factor A. Which simple main effects to test (A or B, but not both) depends on which one is of interest to the researcher. For a *balanced* (A x B) ANOVA design, with n observation in each (A x B) cell, we can compare any two levels of factor A (say, A_j and A_m) at a given level of factor B (say, B_k) by using the following t-statistic:

$$t = \frac{\bar{Y}_{jk} - \bar{Y}_{mk}}{\sqrt{\frac{2\mathrm{MS_W}}{n}}}, \qquad \textbf{(15.27)}$$

where $\bar{Y}_{jk}$ is the mean for the cell (A_j, B_k), located at the "intersection" of levels A_j and B_k, $\bar{Y}_{mk}$ is the mean for the cell (A_m, B_k), located at the "intersection" of levels A_m and B_k, and $\mathrm{MS_W}$ is the *mean squares within* (see Formulas 15.21).

15.1.11 Using SPSS for Two-factor ANOVA

This section addresses the use of SPSS for two-factor ANOVA and presentation of the results in APA format, along with some technical comments.

EXAMPLE 15.5 This example illustrates how to use SPSS for the two-factor ANOVA in Example 15.4, with Tukey post hoc comparisons among the three ethnic groups and graphical representation of the interaction between gender and ethnicity. APA-style summaries and discussion of results are also provided. The SPSS data layout is presented in Figure 15.1 (right panel), with the values of the dependent variable (SCORE = math gain score) and the coding values for the two factors: *Gender* (0 = Female, 1 = Male) and *Ethnicity* (1 = Caucasian, 2 = African-American, 3 = Hispanic). The SPSS steps used for this illustration are:

1. Click **Analyze**, click **General Linear Model**, and click **Univariate**.
2. Click **SCORE**, and click ► to move it into the box
3. Click **Gender**, click **Ethnicity**, and click ► to move them into the box **Fixed Factor(s).**
4. Click **Options**, check the boxes **Descriptive statistics**, **Estimates of effect size**, and **Homogeneity tests**, and then click **Continue**.
5. Click **Post Hoc**, click **Ethnicity**, and click ► to move it into the box **Post Hoc Tests for**
6. Check the **Tukey** box and click **Continue**.
7. Click **Plots**, click **Ethnicity**, and click ►to move it into the box **Horizontal Axis**.
8. Then click **Gender**, and click ►to move it into the box **Separate Lines,** and click **Add** [make sure that the interaction term ***Ethnicity*Gender*** appears in the box **Plots**.
9. Click **OK**.

The resulting SPSS output is provided in Figure 15.5. The **Levene's Test of Equality of Variances** table is not shown because it did not provide a test value [due to the extremely small sample size per cell, *n* = 2]. The descriptive statistics (means and standard deviations) and the interaction plot are not shown in the SPSS output either, but they are both provided (in APA format) in Table 15.1 and Figure 15.6, respectively.

Just for comparison, we can see first that the "boxed" values in the **Tests of Between-Subjects Effects** table are exactly the same as those obtained through manual computations in Example 15.2 (for SS_A, SS_B, SS_{AB}, SS_W, SS_T), Example 15.3 (for MS_A, MS_B, MS_{AB}, MS_W), Example 15.4 (for F_A, F_B, F_{AB}, and their degrees of freedom), and in this section (for $p\eta^2_A$, $p\eta^2_B$, and $p\eta^2_{AB}$) [factor A = *Gender*, factor B = *Ethnicity*].

The values of the omnibus effect size, η^2, for the main effect of factor A (η^2_A), main effect of factor B (η^2_B), and the interaction A x B (η^2_{AB}), are not provided with the SPSS output, but they can be easily computed using Formula 15.24 for the SS values provided with the SPSS output. As shown in the previous section for these data, we have: η^2_A = .193, η^2_B = .400, and η^2_{AB} = .271. Recall that Cohen's guidelines for the magnitude of an effect size (small, medium, large) apply for *eta squared* (η^2), but not for *partial eta squared* ($p\eta^2$) estimates. Likewise, *omega squared* (ω^2), the effect size adjusted for the population, can be computed with Formula 15.26 using the values for the components in this formula provided with the SPSS output. Specifically, as shown in the previous section, we have ω^2_A = .050, ω^2_B = .113, and ω^2_{AB} = .00.

The results in the SPSS output, including those not shown in Figure 15.5, are summarized (in APA style) in Tables 15.1, 15.2, 15.3, and Figure 15.6. Specifically, Table 15.1 provides the means and standard deviations of the groups (by gender and ethnicity) on the dependent variable (math gain score). Table 15.2 summarizes the results in the **Tests of Between-Subjects Tests** table, whereas Table 15.3 summarizes the results in the **Multiple Comparisons** table included in Figure 15.5. The interaction between gender and ethnicity is depicted in Figure 15.6.

In this example we asked whether there were gender differences in the dependent variable (math gain score) and whether such differences may vary across the three ethnic groups. The results in Table 15.2 indicate that there is a statistically significant main effect for gender, $F(1, 6) = 8.53$, $p = .03$, $p\eta^2 = .59$, a statistically significant main effect for ethnicity, $F(2, 6) = 8.84$, $p = .02$, $p\eta^2 = .75$, and a statistically significant interaction between gender and ethnicity, $F(2, 6) = 6.00$, $p = .04$, $p\eta^2 = .67$, at the .05 level of significance. As mentioned previously, when there is a statistically significant interaction, the main effects are of little interest regardless of their statistical significance. They are interpreted here only for the sake of illustration.

The *partial eta squared* ($p\eta^2$) measure of effect size for gender (.59) indicates that 59 percent of the students' differences in math gain scores is accounted for by gender differences, controlling for the effects of ethnicity and the interaction between gender and ethnicity. Likewise, the value of $p\eta^2$ for ethnicity (.75) shows that 75 percent of the differences in math gain scores are accounted for by differences among the ethnic groups, controlling for the effects of gender and the interaction between gender and ethnicity. Also, $p\eta^2$= .67 for the interaction effect size shows that the interaction between gender and ethnicity accounts for 67 percent of the differences in math gain scores controlling for the effects of gender and ethnicity.

Figure 15.5 *Selected SPSS output for the two-factor (Gender x Ethnicity) ANOVA*

Tests of Between-Subjects Effects

Dependent Variable: SCORE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	484.000[a]	5	96.800	7.642	.014	.864
Intercept	1728.000	1	1728.000	136.421	.000	.958
Gender	108.000	1	108.000	8.526	.027	.587
Ethnicity	224.000	2	112.000	8.842	.016	.747
Gender * Ethnicity	152.000	2	76.000	6.000	.037	.667
Error	76.000	6	12.667			
Total	2288.000	12				
Corrected Total	560.000	11				

SS_W SS_T SS_A SS_B SS_{AB} MS_A MS_{AB} MS_B MS_W F_A F_{AB} F_B p-values $p\eta^2_A$ $p\eta^2_{AB}$ $p\eta^2_B$

Multiple Comparisons

Dependent Variable: SCORE
Tukey HSD

(I) Ethnicity	(J) Ethnicity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Caucasian	African-American	-2.00	2.517	.720	-9.72	5.72
	Hispanic	-10.00*	2.517	.017	-17.72	-2.28
African-American	Caucasian	2.00	2.517	.720	-5.72	9.72
	Hispanic	-8.00*	2.517	.044	-15.72	-.28
Hispanic	Caucasian	10.00*	2.517	.017	2.28	17.72
	African-American	8.00*	2.517	.044	.28	15.72

Based on observed means.

*. The mean difference is significant at the .05 level.

Table 15.1

Means and Standard Deviations for Math Gain Score by Gender and Ethnicity

	Female			Male			Total		
Ethnicity	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Caucasian	2	9.00	2.83	2	7.00	4.24	4	8.00	3.16
African-American	2	10.00	2.83	2	10.00	1.41	4	10.00	1.83
Hispanic	2	26.00	2.83	2	10.00	5.66	4	18.00	9.93

Table 15.2

Analysis of Variance for Math Performance

Source	*df*	*F*	$p\eta^2$	*p*
Gender (G)	1	8.53	0.59	0.03
Ethnicity (E)	2	8.84	0.75	0.02
G X E	2	6.00	0.67	0.04
S within group error	6	(12.67)		

Note. The value enclosed in parentheses is the *mean square error* (MS_W). S = subjects.

Table 15.3

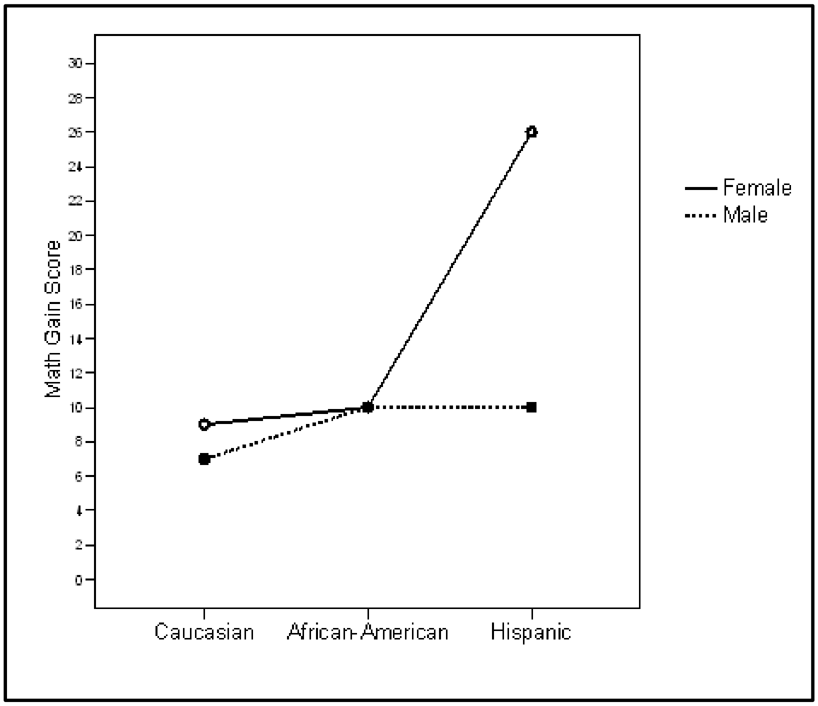
Multiple Comparisons for Math Gain Score Among Ethnic Groups

Ethnic Groups	*ΔM*	SE*ΔM*	95% CI for *ΔM*	
Caucasian — African-American	2.00	2.52	−9.72	5.72
Caucasian — Hispanic	10.00*	2.52	−17.72	−2.28
African-American — Hispanic	−8.00*	2.52	−15.72	−0.28

Note. Δ*M* = Mean difference. SEΔ*M* = Standard error of Δ*M*.

**p* < .05.

Figure 15.6 *Interaction between gender and ethnicity on math gain score*



While pη^2provides useful effect-size information, it cannot be used to compare effect sizes across different effects because the denominator in the formula for pη^2 varies across the effects (see Formula 15.25). For the same reason, Cohen's guidelines for interpreting the magnitude of effect sizes (.01 = small, .06 = *medium*, and .14 = *large*) do not apply for pη^2 (Cohen, 1988). Comparison of effect sizes across different effects can be achieved by reporting the omnibus *eta squared* (η^2) statistic. For the data in this example, the computations in Section 15.1.8 show that there are large effect sizes for gender (η^2 = .193), ethnicity (η^2 = .400), and the interaction between gender and ethnicity (η^2 = .271). Clearly, the largest effect size is associated with the effect of ethnicity (.400), which indicates that 40 percent of the differences in math gain scores are accounted for by gain score differences among the three ethnic groups (Caucasian, African-American, and Hispanic).

Given the statistically significant main effect for ethnicity, the Tukey post hoc method of multiple comparisons was used to determine which ethnic groups differ in math gain scores. The results in Table 15.3 indicate that, at the .05 level, there is no statistically significant difference between the Caucasian and African-American groups (p > .05), but there is a statistically significant difference between the Caucasian and Hispanic groups (p < .05), as well as between the African-American and Hispanic groups (p < .05). Specifically, the results provided by the 95 percent confidence interval for the difference between the means of two groups indicate that the Hispanic group outperformed both the Caucasian group (by a difference between 2.28 and 17.72) and the African-American group (by a difference between 0.28 and 15.72) in math gain scores. Evidently, the Hispanic group in the study population of middle school students benefited most from the experimental school program that incorporates bilingual (English-Spanish) interpretation of mathematics concepts and principles.

As noted earlier, the results on main effects are provided here only to illustrate their APA style presentation. When there is a statistically significant interaction, the main effects are of little interest and can even be misleading. The examination of Figure 15.6 provides more refined information about our major interest in this example — whether there are gender differences in math gain scores across ethnic groups. Specifically, while there are no gender differences for the Caucasian and African-American groups, there is a substantial difference between females (M = 26, SD = 2.83, n = 2) and males (M = 10, SD = 5.66, n = 2) for the Hispanic group in favor of the female students. To test the simple main effect of gender for the Hispanic group, we use Formula 15.27 with $n = 2$, $\bar{Y}_{13} = 26$, $\bar{Y}_{23} = 10$, and $MS_W = 12.67$ (see Table 15.2) to compute the t-statistic:

$$t = \frac{\bar{Y}_{13} - \bar{Y}_{23}}{\sqrt{\frac{2MS_W}{n}}} = \frac{26 - 10}{\sqrt{\frac{(2)(12.67)}{2}}} = 4.50.$$

As the t-statistic (4.50) exceeds the critical t-value (2.447), at the .05 level and with $df = 6$ (see Table A-2), we can conclude that there is a statistically significant simple effect of gender in math gain score (in favor of females) for the Hispanic group of students. This, however, is not the case for the other two ethnic groups. Specifically, the t-statistics for the Caucasian and African-America groups are $t = 0.56$ and $t = 0.00$, respectively, which do not exceed the critical t-value (2.447). The results indicate that the Hispanic female students benefit most from a school program that incorporates bilingual (English-Spanish) interpretation of mathematics concepts and principles. The differences in math gain scores among all other groups—Hispanic (males), Caucasian (males and females), and African-American (males and females)—are negligible.

15.2 Three-factor ANOVA

The logic of two-factor ANOVA carries over directly into ANOVA with three or more factors. Suppose a researcher wants to study whether the motivation for academic achievement of high school students varies according to their gender, ethnicity, and socio-economic status. This question can be addressed by using ANOVA with three factors (gender, ethnicity, and socio-economic status) and the dependent variable "motivation for academic achievement." For concreteness, suppose the three factors are: A = *gender* (male, female), B = *ethnicity* (Caucasian, African-American, Hispanic, Asian), and C = *socio-economic status*, as measured by family income (low, medium, and high). The notation 2 x 4 x 3 ANOVA indicates that this is a three-factor ANOVA, where factor A has 2 levels, factor B has 4 levels, and factor C has 3 levels.

The notations with the two-factor (A x B) AVOVA, used in the previous section, are extended with the three-factor (A x B x C) ANOVA to take into account the presence of a third factor, C. Specifically,

μ_{jkl} = mean of the cell (j, k, l) [the "intersection" of level j of factor A, level k of factor B, and level l of factor C]

$\mu_{j\cdot\cdot}$ = population mean for level j of factor A across all levels of factors B and C; ($j = 1, 2, \ldots, J$),

$\mu_{\cdot k\cdot}$ = population mean for level k of factor B across all levels of factors A and C; ($k = 1, 2, \ldots, K$),

$\mu_{\cdot\cdot l}$ = population mean for level l of factor C across all levels of factors A and B; ($l = 1, 2, \ldots, L$),

$\mu_{\bullet\bullet\bullet}$ = grand mean (of all observations across all levels of factors A, B, and C);
$\alpha_j = \mu_{j\bullet\bullet} - \mu_{\bullet\bullet\bullet}$ [effect of level *j* in factor A],

$\beta_k = \mu_{\bullet k\bullet} - \mu_{\bullet\bullet\bullet}$ [effect of level *k* in factor B],

$\gamma_l = \mu_{\bullet\bullet l} - \mu_{\bullet\bullet\bullet}$ [effect of level *l* in factor C],

$\alpha\beta_{jk} = \mu_{jk} - (\mu_{\bullet\bullet\bullet} + \alpha_j + \beta_k)$ [interaction term of cell (*j*, *k*)],

$\alpha\gamma_{jl} = \mu_{jl} - (\mu_{\bullet\bullet\bullet} + \alpha_j + \gamma_l)$ [interaction term of cell (*j*, *l*)],

$\beta\gamma_{kl} = \mu_{kl} - (\mu_{\bullet\bullet\bullet} + \beta_k + \gamma_l)$ [interaction term of cell (*k*, *l*)],

$\alpha\beta\gamma_{jkl} = \mu_{jkl} - (\mu_{\bullet\bullet\bullet} + \alpha_j + \beta_k + \gamma_l + \alpha\beta_{jk} + \alpha\gamma_{jl} + \beta\gamma_{kl})$ [interaction term of cell (*j*, *k*, *l*)]

$\varepsilon_{ijkl} = Y_{ijkl} - \mu_{jkl}$ [error term for the score of subject *i* within cell (*j*, *k*, *l*)]

σ_ε^2 = *population error variance* [the within-cell variance of the error terms, ε_{ijkl}].

The **assumptions of normality, homogeneity of variance, and independence of observations** hold for ANOVA with three (or more) factors. With the 2 x 3 x 4 ANOVA, for example, there are 24 cells (i.e., 24 population distributions of scores). Under the assumptions of normality and homogeneity of variance, all 24 population (within-cells) distributions are normal and have equal variances, σ_ε^2 [compare to the 2 x 3 ANOVA case depicted in Figure 15.2.]

Under the **linear model** for the data with three-factor ANOVA, the score of any subject *i* in cell (*j*, *k*, *l*) is a linear sum of the effects, interaction terms, and error term as follows:

$$Y_{ijkl} = \mu_{\bullet\bullet\bullet} + \alpha_j + \beta_k + \gamma_l + \alpha\beta_{jk} + \alpha\gamma_{jl} + \beta\gamma_{kl} + \alpha\beta\gamma_{jkl} + \varepsilon_{ijkl}. \quad \textbf{(15.28)}$$

There are seven null hypotheses testable with the tree-factor ANOVA:

Main effect of factor A, H_{01}: $\mu_{1\bullet\bullet} = \mu_{2\bullet\bullet} = \ldots = \mu_{J\bullet\bullet}$ [or, H_{01}: $\sum \alpha_j^2 = 0$]

Main effect of factor B, H_{02}: $\mu_{\bullet 1\bullet} = \mu_{\bullet 2\bullet} = \ldots = \mu_{\bullet K\bullet}$ [or, H_{02}: $\sum \beta_k^2 = 0$]

Main effect of factor C, H_{03}: $\mu_{\bullet\bullet 1} = \mu_{\bullet\bullet 2} = \ldots = \mu_{\bullet\bullet \mathrm{L}}$ [or, H_{03}: $\sum \gamma_l^2 = 0$]

A x B interaction, H_{04}: There is no A x B interaction [or, H_{04}: $\sum\sum(\alpha\beta_{jk})^2 = 0$]

A x C interaction, H_{05}: There is no A x C interaction [or, H_{05}: $\sum\sum(\alpha\gamma_{jl})^2 = 0$]

A x C interaction, H_{06}: There is no B x C interaction [or, H_{06}: $\sum\sum(\beta\gamma_{kl})^2 = 0$]

A x B x C interaction, H_{07}: There is no A x B x C interaction [or, $\sum\sum\sum(\alpha\beta\gamma_{jkl})^2 = 0$]

The *F*-test for the three null hypotheses in two-factor ANOVA applies to testing the null hypotheses with three-factor ANOVA. The *F*-statistics in this case are:

Main effects: $F_\mathrm{A} = \dfrac{\mathrm{MS_A}}{\mathrm{MS_W}}$ (for H_{01}), $F_\mathrm{B} = \dfrac{\mathrm{MS_B}}{\mathrm{MS_W}}$ (for H_{02}), and $F_\mathrm{C} = \dfrac{\mathrm{MS_C}}{\mathrm{MS_W}}$ (for H_{03}),

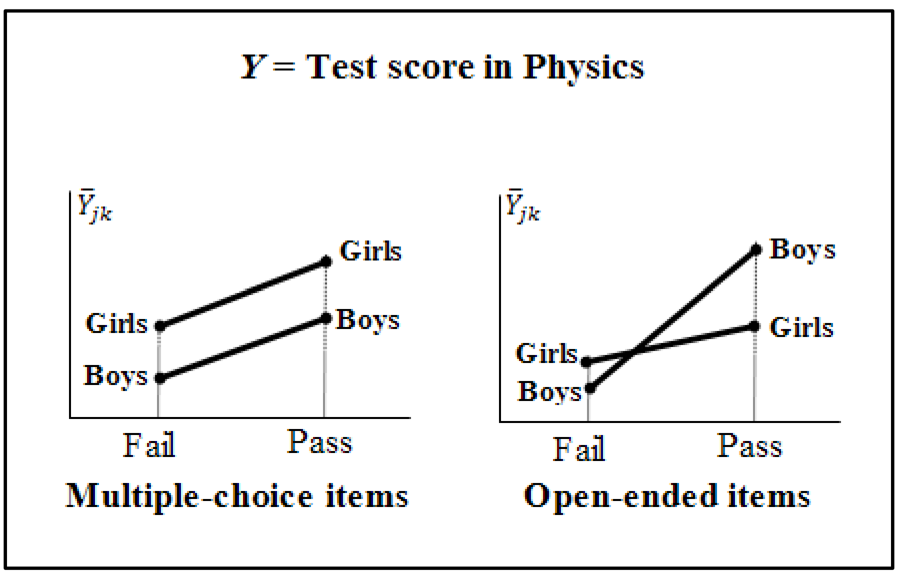
Dual interactions: $F_\mathrm{AB} = \dfrac{\mathrm{MS_{AB}}}{\mathrm{MS_W}}$ (for H_{04}), $F_\mathrm{AC} = \dfrac{\mathrm{MS_{AC}}}{\mathrm{MS_W}}$ (for H_{05}), and $F_\mathrm{BC} = \dfrac{\mathrm{MS_{BC}}}{\mathrm{MS_W}}$ (for H_{06}),

Triple interaction: $F_{\text{ABC}} = \dfrac{\text{MS}_{\text{ABC}}}{\text{MS}_{\text{W}}}$ (for H_{07}).

Evidently, the testing for main effects and dual interactions (between two factors) is the same as with two-factor ANOVA. However, the power of the F-tests with three-factor ANOVA is higher than that with the two-factor ANOVA because the MS_{W} with the former is generally smaller than the MS_{W} with later [Why?]

The new element in the transition from a two-factor ANOVA to a three-factor ANOVA is the "triple" (three-factor) interaction A x B x C. Generally, a **three-factor interaction** occurs when there is an interaction between two factors at some level(s) of the third factor, but this interaction changes (or disappears) at some other level(s) of the third factor. A three-factor interaction in a 2 x 2 x 2 ANOVA is depicted in Figure 15.7. The factors are *gender* (boys, girls), *test performance* (fail, pass), and *test form* — a sample of middle school students was randomly split and each group randomly assigned to take a Physics test in either multiple-choice items (MCI) or open-ended items (OEI) format. The total scores on the test represent the dependent variable, Y. The mean score in Y for boys and girls who failed or passed the test, $\bar{Y}_{jk}$, is presented on the vertical axis, separately for the MCI and OEI test forms. The plots in Figure 15.7 are based on a real data analysis which showed that the three-factor interaction is statistically significant, thus opening the door for interpretation of the interaction plots. As can be seen, girls consistently outperform boys on the MCI test form, but the difference remains the same (in direction and magnitude) for students who failed and for those who passed the test. This trend cannot be generalized across test forms, as the picture for the OEI test form is quite different — for students who failed the test, girls did slightly better than boys, whereas for students who passed the test, boys did much better than girls.

Figure 15.7 *Three-factor interaction among* ***gender****,* ***test performance****, and* ***test form*** *for the test scores of middle school students*

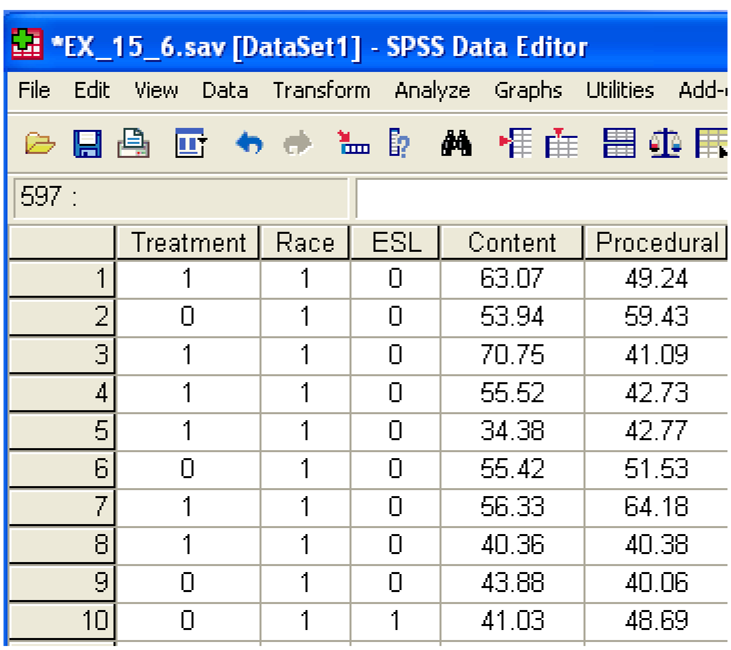


NOTE [15.3] To proceed with the interpretation of interaction plots, there must be statistical evidence for the presence of interaction. For a three-factor interaction, such evidence is provided (or not) by the respective F-test (see F_{ABC} for H_{07}).

EXAMPLE 15.6 This example illustrates how to use SPSS to conduct a three-factor ANOVA and how to interpret the results. The SPSS data file **EX_15_6.sav** is available on the website of this book [http://cehd.gmu.edu/book/dimitrov]. The data consist of 597 observations on the following five variables: **Treatment** (0 = Control, 1 = Experimental), **Race** (1 = White, 2 = Black, 3 = Other), **ESL** = English as a Second Language (0 = No, 1 = Yes), **Content** = Scores on a *content knowledge* test (on a T-scale, *Mean* = 50, *SD* = 10), and **Procedural** = Scores on a *procedural knowledge* test (also on a T-scale). The first 10 (out of 597) observations are shown in Figure 15.8. The three factors are Treatment, Race, and ESL, so a 2 x 3 x 2 ANOVA is employed with *content knowledge* as the dependent variable. [The use of *procedural knowledge* as the dependent variable to conduct a three-way ANOVA is requested with study question 7 at the end of this chapter.]

The research question is whether an experimental program in teaching science makes a difference in the students' content knowledge, taking into account the students' race and whether English is their second language (ESL). Gender is not considered here because preliminary analyses showed that gender was not involved in any statistically significant main or interaction effects. The null hypotheses with the 2 x 3 x 2 ANOVA in this case relate to main effect of Treatment (T), main effect of Race (R), main effect of ESL, three pairwise interactions between factors (T x R, T x ESL, R x ESL), and a three-factor interaction, T x R x ESL.

Figure 15.8 *SPSS data file* **EX_15_6.sav** *— the first 10 (out of 597) observations*



	Treatment	Race	ESL	Content	Procedural
1	1	1	0	63.07	49.24
2	0	1	0	53.94	59.43
3	1	1	0	70.75	41.09
4	1	1	0	55.52	42.73
5	1	1	0	34.38	42.77
6	0	1	0	55.42	51.53
7	1	1	0	56.33	64.18
8	1	1	0	40.36	40.38
9	0	1	0	43.88	40.06
10	0	1	1	41.03	48.69

Following the SPSS steps described in Example 15.5 for a two-factor ANOVA, but this time using three fixed factors (**Treatment, Race**, **ESL**) and **Content** as a dependent variable, we obtain the SPSS output provided in Figure 15.9. For space consideration, the results are not summarized in APA style, but this can be done following the illustration in Tables 15.1, 15.2, and 15.3 for two-factor ANOVA. The three-factor interaction plots provided with the SPSS output, edited for APA format, are shown in Figure 15.10.

Figure 15.9 *Selected SPSS output for the three-factor (Treatment* x *Race* x *ESL) ANOVA*

==

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Content knowledge

F	df1	df2	Sig.
.690	11	585	.749

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+Treatment+Race+ESL+Treatment * Race+Treatment * ESL+Race * ESL+Treatment * Race * ESL

Tests of Between-Subjects Effects

Dependent Variable: Content knowledge

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3318.698[a]	11	301.700	3.136	.000	.056
Intercept	631669.899	1	631669.899	6565.713	.000	.918
Treatment	659.372	1	659.372	6.854	.009	.012
Race	1562.842	2	781.421	8.122	.000	.027
ESL	925.942	1	925.942	9.624	.002	.016
Treatment * Race	187.207	2	93.603	.973	.379	.003
Treatment * ESL	262.617	1	262.617	2.730	.099	.005
Race * ESL	1012.715	2	506.357	5.263	.005	.018
Treatment * Race * ESL	622.999	2	311.500	3.238	.040	.011
Error	56281.302	585	96.207			
Total	1552100.000	597				
Corrected Total	59600.000	596				

a. R Squared = .056 (Adjusted R Squared = .038)

Multiple Comparisons

Dependent Variable: Content knowledge
Tukey HSD

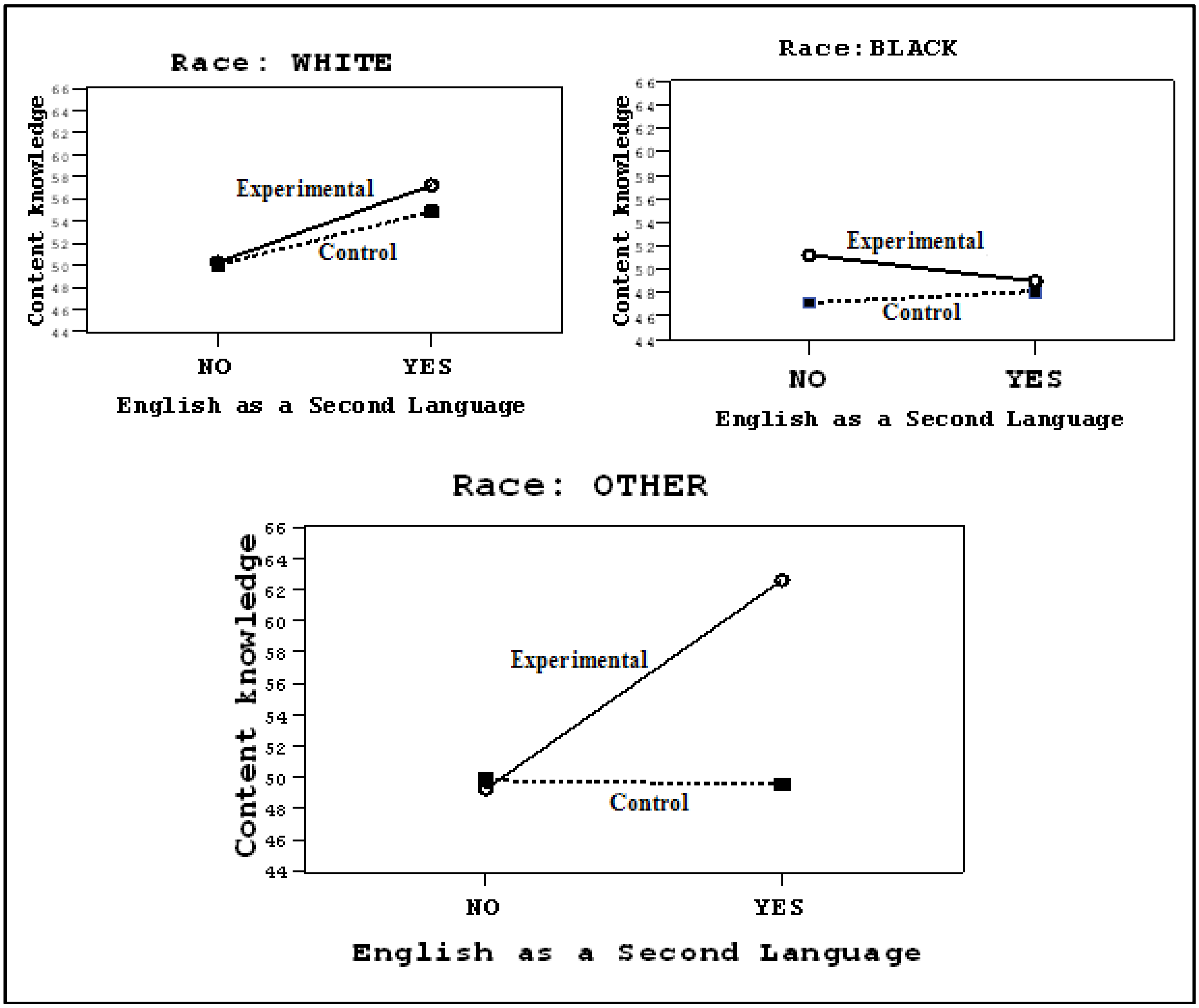
(I) Race	(J) Race	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
WHITE	BLACK	2.2959*	.86199	.022	.2704	4.3213
	OTHER	.0979	1.32339	.997	-3.0117	3.2074
BLACK	WHITE	-2.2959*	.86199	.022	-4.3213	-.2704
	OTHER	-2.1980	1.36511	.242	-5.4056	1.0096
OTHER	WHITE	-.0979	1.32339	.997	-3.2074	3.0117
	BLACK	2.1980	1.36511	.242	-1.0096	5.4056

Based on observed means.

*. The mean difference is significant at the .05 level.

==

Figure 15.10
Three-factor (Treatment × Race × ESL) Interaction for Test Scores on Content Knowledge



The results from the Levene's test show that the homogeneity of variance assumption is met for the data in this example, $F(11, 585) = 0.69$, $p = .75$. For space consideration, the means and standard deviations are not shown in Figure 15.9, but they are reported here. The results from the tests of between-subjects effects indicate that there is a statistically significant main effect of Treatment, $F(1, 585) = 6.85$, $p < .01$, $p\eta^2 = .01$, with higher scores for the experimental group ($M = 53.25$, $SD = 1.00$, $n = 198$) compared to the control group ($M = 49.92$, $SD = 0.79$, $n = 399$). There was also a statistically significant main effect of ESL, $F(2, 585) = 9.62$, $p < .01$, $p\eta^2 = .02$, with higher scores for the students with English as a second language ($M = 53.56$, $SD = 1.12$, $n = 159$) than those with native English speakers ($M = 49.61$, $SD = 0.61$, $n = 438$). The main effect for Race was also statistically significant, $F(2, 585) = 8.12$, $p < .001$, $p\eta^2 = .03$. The results from the Tukey post hoc test for Race, reported in Figure 15.9, show that the White group outperformed the Black group ($p < .05$) by a difference that varies between 0.27 and 4.32 units on the test scale, but there were no other differences among the three racial groups.

Although the three main effects are statistically significant, they are of little interest, as there is a statistically significant interaction between Race and ESL, $F(2, 585) = 5.26$, $p < .01$), $\text{p}\eta^2 = .02$, as well as a statistically significant three-factor interaction, Treatment x Race x ESL, $F(2, 585) = 3.24$, $p < .05$, $\text{p}\eta^2 = .01$. Recall that the main research question in this example is whether an experimental program in teaching science makes a difference in the students' content knowledge, taking into account the students' race and whether English is their second language. In the presence of an interaction, the statistically significant main effect of Treatment cannot be generalized across the levels of Race and ESL. Therefore, the difference between the experimental and control groups must be examined in the context of the three-factor interaction depicted in Figure 15.10.

The three-factor interaction plots in Figure 15.10 were obtained by running a two-factor (Treatment x ESL) ANOVA separately for each level of the factor Race. This is achieved by first "splitting" the SPSS file by Race [click **Data**, click **Split File**, select **Organize output by groups**, select **Race**, and click ► to move it into the box **Groups Based on].** Then run the two-factor (Treatment x ESL) ANOVA as shown with the SPSS steps in Example 15.5.

The results showed that the interaction between Treatment and ESL was not statistically significant for the White and Black racial groups, but it was statistically significant for the "Other" group ($p < .05$). Clearly, the largest difference between the experimental and control groups is for students in the "Other" group for whom English is a second language (ESL = 1). This result, however, must be interpreted with caution, given the relatively small number of observations in the "Other" group ($n = 67$) compared to the number of observations in White ($n = 305$) and Black ($n = 225$) groups. Moreover, the number of observations by Race becomes even smaller when distributed by cells across the levels of the other two factors, Treatment and ESL.

In addition, as the data in this example (**EX_15_6.sav**) do not produce a balanced three-factor ANOVA design, we cannot test for simple main effects using the t-test calculated by Formula 15.27. Thus, assuming valid data prerequisites for the results in this example, we can conditionally summarize that the experimental group outperforms the control group, but while the treatment effect is negligible for students from the White and Black racial groups, regardless of their ESL status, it is well pronounced in favor of the students from the "Other" racial group for whom English is a second language.

NOTE [15.4] When the ANOVA design is not balanced, we cannot use Formula 15.27 (or other formulas that assume balanced ANOVA) to test for simple main effects. Procedures for simple main effects with unbalanced ANOVA are beyond the scope of this chapter, but more on this topic can be found on the website for this book, http://cehd.gmu.edu/book/dimitrov.

15.3 Summary

Using ANOVA with two or more factors allows for (a) investigating the role of different sources in accounting for the variance in the dependent variable, (b) detecting and interpreting interactions between factors, and (c) increasing the power of tests for differences among factor levels and interactions among factors.

- **Assumptions in ANOVA**
  The three primary assumptions in one-factor ANOVA, *independence*, *normality*, and *homogeneity of variance*, remain in place for ANOVA with two or more factors. The difference is that, with adding a new factor in ANOVA, the population distributions (subjects submitted to the same ANOVA conditions) become more homogeneous and, as a result, the population error variance, σ_ε^2, becomes smaller (e.g., compare Figures 14.1 and 15.2).

- **Effects in two-factor ANOVA**
  There are three types of effects in two-factor (A x B) ANOVA, α_j — *row effects* (for the levels of factor A), β_k — *column effects* (for the levels of factor B), and $\alpha\beta_{jk}$— *cell effects* (interaction terms), defined with Equations 15.4, 15.5, and 15.6, respectively.

- **Null hypotheses in two-factor ANOVA**
  There are three testable null hypotheses with two-factor ANOVA, with *J* levels in Factor A and *K* levels in Factor B,:

  main effect of A, H_{01}: $\mu_{1\bullet} = \mu_{2\bullet} = \ldots = \mu_{J\bullet}$ [or, H_{01}: $\sum \alpha_j^2 = 0$],
  main effect of B, H_{02}: $\mu_{\bullet 1} = \mu_{\bullet 2} = \ldots = \mu_{\bullet K}$ [or, H_{02}: $\sum \beta_k^2 = 0$], and
  interaction A x B, H_{03}: There is NO interaction between factors A and B
  [or, H_{03}: $\sum\sum(\alpha\beta_{jk})^2 = 0$.]

- **Linear model for the data in two-factor ANOVA**
  Under the linear model assumption in two-factor ANOVA, any observed score can be represented as a sum of the total effect ($\mu_{\bullet\bullet}$), row effect (α_j), column effect (β_k), cell effect ($\alpha\beta_{jk}$), and error "effect" (ε_{ijk}) for this observation:

  $$Y_{ijk} = \mu_{\bullet\bullet} + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk}.$$

- **Testing the null hypotheses in two-factor ANOVA**
  The null hypotheses in two-factor ANOVA (H_{01}, H_{02}, and H_{03}) are tested by using the *F*-statistics: $F_A = MS_A/MS_W$, $F_B = MS_B/MS_W$, and $F_{AB} = MS_{AB}/MS_W$, respectively.

- **Measures of effect size in two-factor ANOVA**
  Eta squared (η^2) measures an omnibus effect size of an "effect" (A, B, or A x B) by indicating what proportion of the total variability in the dependent variable (SS_T) is accounted for by this effect (see Formula 15.24). *Partial eta squared* ($p\eta^2$) indicates the proportion of the variability in the dependent variable, *Y*, accounted for by a given effect after "partialling out" (controlling for) the contribution of all other effects (see Formula 15.25). Omega squared (ω^2) is an adjustment of η^2 for the population (see Formula 15.26). The Cohen's guidelines for interpreting

the magnitude of the effect size η^2 (*small*: η^2 = .01, *medium*: η^2= .06, and *large*: η^2 = .14) do not apply for partial eta squared, $p\eta^2$ (Cohen, 1988).

- **Types of interaction in two-factor ANOVA**

  There are two types of interaction that may occur in two-factor ANOVA — ordinal or disordinal interaction. *Ordinal interaction* is when the order of the mean scores for the levels in one factor is the same across the levels of other factor. *Disordinal interaction* is when the order of mean scores for the levels in one factor is not the same across the levels of other factor (see Figure 15.4).

- **Testing for simple main effects in two-factor ANOVA**

  If the (A x B) interaction in a two-factor ANOVA is statistically significant, *testing for simple main effects* of factor A is performed by testing the difference between any two levels of factor A across the levels of factor B. With a balanced ANOVA design, the testing for simple main effects can be performed by using the *t*-test statistic in Formula 15.27.

- **Three-factor ANOVA**

  The logic of two-factor ANOVA carries over directly into ANOVA with three or more factors. The main difference is that, with a three factor (A x B x C) ANOVA, there are seven testable null hypotheses — main effects of A, B, and C, interactions between two factors (A x B, A x C, and B x C), and a three-factor interaction, A x B x C. A statistically significant A x B x C interaction means that the interaction pattern between any two factors is not the same across the levels of the third factor (e.g., see Figure 15.10).

15.4 Study Questions

1. Table 15.4.1 provides the data layout for a two-factor ANOVA with factors Gender (1 = Female, 2 = Male) and Treatment (1 = Control, 2 = Experimental) with two observations within each cell (n = 2). Using these data, (a) compute the row means, the column means, and the grand mean, and (b) compute the row effects, column effects, and interaction terms. [*Hint*: see Formula 15.4-15.7.]

   Table 15.4.1

	Control	Experimental
Female	1, 3	40, 16
Male	2, 6	30, 14

2. For the two-factor ANOVA data in Table 15.4.1, compute (a) the sum of squares (SS_A, SS_B, and SS_{AB}) and (b) the mean squares (MS_A, MS_B, MS_{AB}, and MS_W), where factor A = *Gender* and factor B = *Treatment*. [*Hint*: see Sections 15.1.5 and 15.1.6]

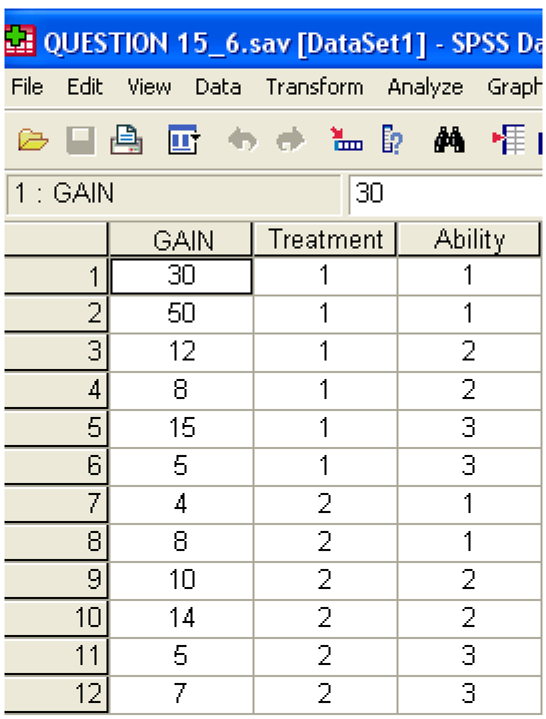
3. Given the data layout in Table 15.4.1, (a) formulate the null hypotheses for the two-factor ANOVA and (b) test the null hypotheses by computing the *F*-statistics (F_A, F_B, and F_{AB}), where factor A = *Gender* and factor B = *Treatment*. [*Hint*: see Section 15.1.7]

4. For the two-factor ANOVA data in Table 15.4.1, compute the effect sizes (a) eta squared (η^2), (b) partial eta squared ($p\eta^2$), and (c) omega squared (ω^2) for A, B, and AB, where factor A = *Gender* and factor B = *Treatment*. [Hint: see Section 15.1.8]

5. For the two-factor ANOVA data in Table 15.4.1, use SPSS to analyze the data and then present and interpret the results (in APA format). [*Hint*: see Example 15.5]

6. Table 15.4.2 provides the data layout for a two-factor ANOVA. The dependent variable GAIN indicates the "gain" score of students in reading under two treatment conditions labeled Treatment (1 = Control, 2 = Experimental), which represent two different approaches to teach reading. The variable Ability has three levels (1 = Low, 2 = Middle, and 3 = High) indicating the student's reading ability prior to "treatment." The research question is whether the treatment conditions make a difference in the students' gain score on reading and does the difference depend on the prior ability level of the students in reading. Use SPSS to perform the two-factor (Treatment x Ability) ANOVA and then present and interpret the results in APA format. Make sure to test for simple main effects in case there is a statistically significant interaction between Treatment and Ability [*Hint*: see Section 15.1.10].

Table 15.4.2


QUESTION 15_6.sav [DataSet1] - SPSS Da

File Edit View Data Transform Analyze Graph

1 : GAIN 30

	GAIN	Treatment	Ability
1	30	1	1
2	50	1	1
3	12	1	2
4	8	1	2
5	15	1	3
6	5	1	3
7	4	2	1
8	8	2	1
9	10	2	2
10	14	2	2
11	5	2	3
12	7	2	3

7. Using the SPSS data file **EX_15_6.sav** [http://cehd.gmu.edu/book/dimitrov], perform the three-factor (Treatment x Race x ESL) ANOVA and interpret the results, as shown in Example 15.6, but this time use the variable **Procedural** (the students' scores on procedural knowledge) as the dependent variable. Present the results in APA style following the format in Tables 15.1, 15.2, and 15.3 adapted for the three-factor ANOVA in this case.