Chapter 9

FACTORIAL ANALYSIS OF VARIANCE

When researchers have more than two groups to compare, they use analysis of variance, which is commonly called "ANOVA." When they have more than one independent variable as well, the comparison of means is called "*factorial* analysis of variance."[1] This chapter explains the ways researchers may apply this versatile tool. The steps involved in completing

[1]It is important not to confuse factorial analysis of variance with factor analysis, which is a method to examine intercorrelations among variables to determine the number and nature of any underlying dimensions.

any factorial ANOVA study will be examined. Then, the use of standard factorial ANOVA will be followed by applications for random, mixed, and repeated measures applications.

DOING A STUDY THAT INVOLVES MORE THAN ONE INDEPENDENT VARIABLE

When researchers explore differences among more than two means from two or more variables, several processes are required as minimums. Five major steps are most prominent.

1. *Stating a Hypothesis for Factorial ANOVA.* A research hypothesis for a factorial analysis of variance deals with the differences among means. A hypothesis where ANOVA is an appropriate test has one or more independent variables measured at the nominal level and a dependent variable measured at the interval level. For instance, a researcher might wish to compare the communicator competence ratings of students who receive a combination of case study readings and classroom instruction on responsive listening. The students could be randomly assigned to receive (or not to receive) classroom instruction in responsive listening. At the same time, they could be asked to read sets of interpersonal communication case studies involving people in low, moderate, or high intimacy relationships. The researcher might state hypotheses for each independent variable (degree of classroom instruction in responsive listening and relationship intimacy levels of case studies read).

   - The researcher might hypothesize that students receiving classroom instruction in responsive listening have higher communicator competence scores than students not receiving classroom instruction in responsive listening. The hypothesis would be

     $$H\text{: } \mu_{\text{students receiving classroom instruction in responsive listening}} > \mu_{\text{students not receiving classroom instruction in responsive listening}}.$$

     The null hypothesis tested statistically would be

     $$H_0\text{: } \mu_{\text{students receiving classroom instruction in responsive listening}} = \mu_{\text{students not receiving classroom instruction in responsive listening}}.$$

     Because there are only two levels for this variable, if the null hypothesis were rejected and the means were as predicted, the research hypothesis would be supported.

   - For the relationship intimacy levels of the case studies variable, the researcher might predict that students reading case studies with high relationship intimacy would have significantly higher mean communicator competence scores than students reading case studies with moderate relationship intimacy, who would also have higher mean communicator competence ratings than students reading case studies with low relationship intimacy. The hypothesis would be

     $$H\text{: } \mu_{\text{reading high relationship intimacy case studies}} > \mu_{\text{reading moderate relationship intimacy case studies}} > \mu_{\text{reading low relationship intimacy case studies}}.$$

     Yet, analysis of variance is an omnibus test of the null hypothesis that there simply is no difference among the means. The null hypothesis would be

     $$H_0\text{: } \mu_{\text{reading high relationship intimacy case studies}} = \mu_{\text{reading moderate relationship intimacy case studies}} = \mu_{\text{reading low relationship intimacy case studies}}.$$

Special Discussion 9.1

Special Statistical Notation for Factorial Designs

Though there is no rule requiring it, most researchers dealing with factorial experiments use a fairly consistent notation form. The letter *j* is used to represent a variable represented on the rows. The letter *k* is employed to represent the variable on the columns. As typically is the case, *i* is used to reference an instance of a particular data point. Thus, a researcher might state $\sum X_{ijk}$ to indicate a desire to sum each instance of a score from each row and column cell.

Dots also are used in factorial experiments to represent a collection of events that go all the way through the total. For instance, a researcher would use the notation $\overline{X}_{1.}$ to refer to the mean of the first row computed from all the instances of data in the row.* Similarly, researchers who see $\overline{X}_{.2}$ know that the symbol refers to the mean of the second column and is computed from all the data through the column. Similarly, $\overline{X}_{..}$ may refer to the grand mean, but the symbols $\overline{\overline{X}}$ or $\overline{\overline{\overline{X}}}$ often are used instead.

*Though using $\overline{X}$ to indicate the mean is the most common, many research reports use *M* to symbolize means. This notation may be complicated by the fact that some communication researchers also use *M* to refer to the number of messages included in a research study.

Rejecting the null hypothesis could mean many things. Therefore, researchers who wish to speculate that there is a difference between one group and a collection of others must follow up with specific comparisons involving contrasts between specific combinations of means. Researchers using ANOVA also can present hypotheses that deal with combinations of variable levels. Such research hypotheses could look something like:

H: $\mu_{\text{highly intelligent}} > (\mu_{\text{moderately intelligent}} + \mu_{\text{lowly intelligent}})/2$,

for which the null hypothesis would be

H_0: $\mu_{\text{highly intelligent}} = (\mu_{\text{moderately intelligent}} + \mu_{\text{lowly intelligent}})/2$.

Obviously, researchers would need to use tools to make specific comparisons in addition to the general ANOVA test.

- Hypotheses about interactions also may posit specific combinations of levels of two variables that would be distinguished from others. For instance, researchers might speculate that highly intelligent women would have higher communication competence than any other combination of levels. Such specific hypotheses would have to be followed by individual contrasts.

2. *The Selection of Levels and Conditions.* Researchers must select variable levels with some care. When researchers can justify selecting variable ranges to reflect the scope of the independent variable of interest, then researchers can state that they are concerned with **fixed effects**. These levels generally are expected to cover the range in which the variable operates normally. If the researcher has a primary variable of interest, a hypothesis usually will accompany the selection of the variable and its levels. Sometimes researchers attempt to identify—roughly speaking—a random selection of variable levels to determine if the variable produces effects in general. Researchers who take this approach are said to explore **random effects**.

3. *Inclusion of Additional Variable Suspected of Interacting With Other Predictors.* The first set of variables may not tell the whole story. In addition to considering adding other variables that would produce important effects one at a time, researchers often include other variables that might be involved in interactions with the primary variables of interest.

4. *Organizing the Conditions and Coding Data.* Researchers must prepare study conditions, collect data, and score measures. For instance, researchers might design an experiment composed of two independent variables, source credibility (low and high) and use of humor in a message (without jokes and with jokes). Researchers would need to prepare all possible combinations of these variable levels. In this case, every participant who happened to be exposed to a message without jokes could receive a code of 1 for a variable called "Humor." Those exposed to a message with jokes could receive a code of 2 for the "Humor" variable. Similarly, those exposed to the message attributed to a source with little credibility could receive 1 for a variable called "Credibility," and those exposed to the message attributed to a highly credible source could receive 2 for the variable.[2] Table 9.1 illustrates the possible interactions of these two variables.

Table 9.1

		Humor	
		Without Jokes	*With Jokes*
Source Credibility	Low		
	High		

5. *Testing Assumptions.* When completing factorial analysis of variance, researchers regularly test assumptions underlying the statistic itself. These tests affect the how analyses may be completed and, hence, will be considered for different forms of ANOVA (fixed, random, mixed, and repeated measures designs) covered in this chapter.

TYPES OF EFFECTS TO TEST

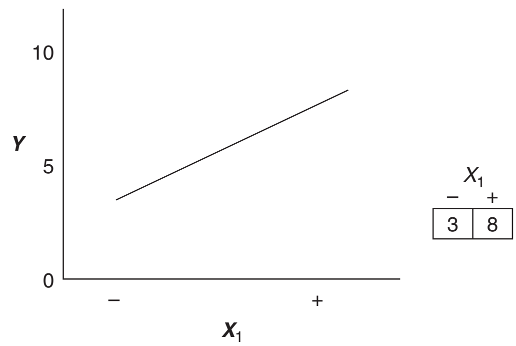
A major reason to complete a factorial analysis of variance study is to identify effects from variables taken separately and in combinations. Each of these effect types will be considered.

Isolating Main Effects

Because factorial designs involve more than one independent variable, researchers naturally care whether these variables produce effects one at a time. Called a **main effect**, this type of

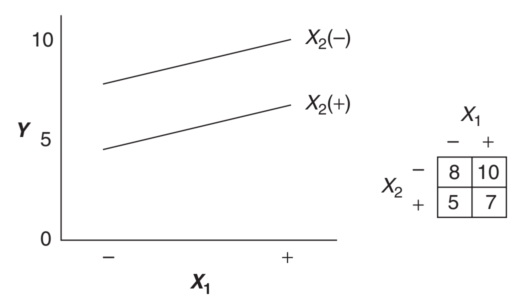
[2]Researchers sometimes find that credibility manipulations are difficult. When made the object of a manipulation check, some "low credibility" conditions turn out to be moderately credible. In some cases, the low credibility induction creates a strong positive violation of expectations (e.g., a high school student talking intelligently about the International Monetary Fund).

result is "the effect of an independent variable uninfluenced by other variables" (Vogt, 2005, p. 183). Because they are main effects, their contributions to total variation in the dependent variable are additive. That is, the factors do not interact, and they produce effects that may be added to other effects to determine the overall impact of variables.

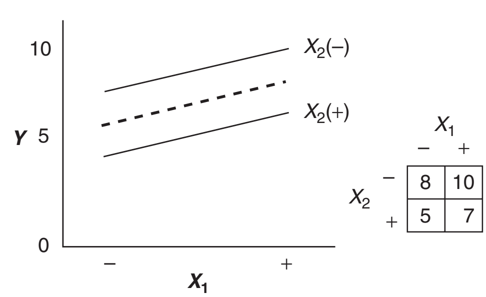


In the figure on the right, the numbers inside the 1 × 2 grid indicate mean scores on the dependent variable, and the identifiers outside the cells show independent variable levels. If only one independent variable is present, as in the case here, the main effect may be identified by looking at a simple line graph, such as shown on the diagram. The dependent variable is indicated on the vertical axis (also known as the *y*-axis or the ordinate). Though bar charts often are used to report research data, line graphs, rather than bar charts, traditionally are employed to identify different sorts of effects on dependent variables.

In this case, the diagram shows that when one moves from the low level of the independent variable (symbolized X_1) to the high level, the dependent variable scores show an upward slope. Indeed, main effects are indicated for the variable on the horizontal axis (also known as the *x*-axis or abscissa) by the existence of line(s) with some slope. A flat line, running parallel to the horizontal axis, would reveal the absence of a main effect. Such a flat line would mean that as one moves from one level to another on the independent variable, there is no change in the dependent variable. If there is a difference in the numbers, there is a main effect. The bigger the difference, the bigger is the effect.

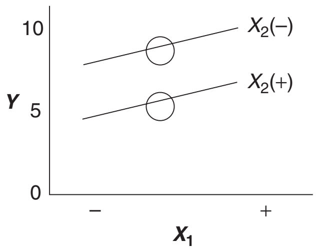


When two independent variables are included, the researcher runs out of available axes on a two-dimensional figure. Thus, separate lines are drawn to represent levels of the second dependent variable, as shown in the diagram on the right. Separate lines have been drawn to show the effects of the levels of variable 2.[3] To reveal whether there is a main effect from the independent variable on the horizontal axis, the researcher looks at the average dependent variable score when the independent variable is at each of its levels (often placing an imaginary dot on these positions). When an imaginary line connecting these means shows slope, a main effect is claimed. In this case, as shown in the diagram on the right, the dotted line shows "imaginary line" and, because it has slope, a main effect exists from independent variable 1.



[3]References to "high" and "low" levels refer only to the levels of the independent variables, not what impact these variables have on the dependent variable.

To detect a main effect from the variable for which separate lines were drawn, the researcher looks at the average of each line and looks for a difference in the dependent variable scores. In short, the researcher looks for space between the middles of the lines. In the case of the example shown here on the left, the middle of each line is indicated by a circle, and there clearly is space between these points. Thus, in this example the researchers would identify a main effect from variable 2.



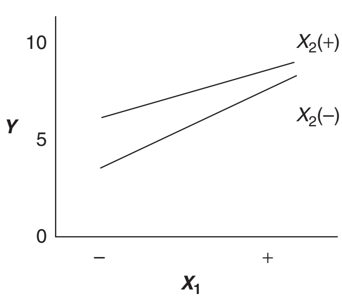
		X_1 −	X_1 +
X_2	−	8	10
	+	5	7

Isolating Interaction Effects

In addition to main effects, researchers often are interested in exploring the effects of special combinations of levels of variables. **Interaction effects** are the "joint effect[s] of two or more independent variables on a dependent variable" (Vogt, 2005, p. 154). Whereas main effects are additive effects, interactions are multiplicative. In other words, the interactions are more than just the result of main effects moving in concert. Instead, they are contributions to overall effects that go beyond the addition of main effects alone. A direct way to identify the nature of interactions is to look at the diagrams of means. In general, interaction effects are revealed when the lines are *nonparallel.* In fact, they may be so nonparallel that they cross (but it is not necessary for lines to cross for an interaction to be present). As we shall see, however, the type of interaction found greatly affects the sorts of interpretations that can be made of the study results.

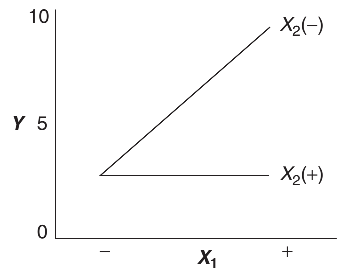
Two types of interactions regularly are identified in research reports. **Uncrossed or ordinal interactions** are dependent interaction effects that are in the same direction as the main effects of the variables involved (Reinard, 2001, p. 439). On the other hand, **crossed or disordinal interactions** are dependent interaction effects that are *not* in the same direction as the main effects of the variables involved (Reinard, 2001, p. 435).

In essence, ordinal interactions indicate the presence of a sort of "bonus effect" when levels of independent variables are put together. The interaction is in the same direction as the main effects, just more (or less) so. The diagram on the left reveals a slight tendency toward nonparallel lines. Obviously, it would be nice to know if such interaction effects contribute variation that is beyond random sampling error alone. Factorial analysis of variance provides such information and will be described shortly. In this example, the main effects reveal lowest dependent variable scores when variable X_1 is low and also when variable X_2 is low. When these two levels are combined, however, the effect goes beyond a simple average of the main effects. Hence, the effect is not additive. By the way, in this case the so-called "bonus effect" is in a negative direction—the dependent variable scores are reduced greatly by the combination of variable levels.



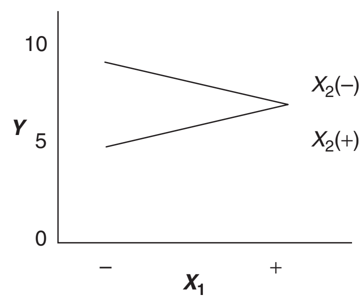
		X_1 −	X_1 +
X_2	−	3	6
	+	8	9

A variation of this type of ordinal interaction involves finding what some researchers call a "magic cell" in which only one combination of conditions produces dependent variable scores that are different from the others. The diagram on the right illustrates such a situation, in which the only time that a different dependent variable score is observed is when independent variable X_2 is set at its low level and independent variable X_1 is set at its high level. It is important to remember, however, that the "high" and "low" levels of independent variables refer only to *their* settings, not the predicted scores on dependent variables.



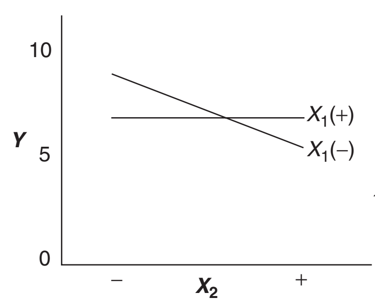
X_2 \ X_1	–	+
–	3	10
+	3	3

Another ordinal interaction form occurs when levels of one variable eliminate the contributions from the other variable. For instance, in the diagram below on the right, when variable X_1 is low, the different levels of variable X_2 influence the dependent variable scores. But when variable X_1 is high, different levels of variable X_2 have no influence on the dependent variable. One might wonder if this interaction is ordinal or disordinal, because the lines touch. The answer is that the interaction is *ordinal* because the lines do not cross. One might speculate that the interaction could be disordinal if the lines were extended further. Yet, interpretations are based on the actual levels and "operating ranges" selected by the researchers. If it is assumed that independent variable levels are set on the basis of some logical reasoning, it does not matter that the lines might cross if somebody else did a different study using different levels of independent variables. In fact, if the levels were simple categories, such as whether the respondents were male of female, it would not make any sense to wonder what a "higher" level of the variable would be. Furthermore, the researcher does not really know that the lines might not have retreated in he opposite direction if the levels of the independent variable had been extended.



X_2 \ X_1	–	+
–	9	7
+	5	7

It is important for researchers to check alternative ways of drawing charts. For instance, if the data from the last example were graphed by switching which variable appears on the horizontal axis, the results would be very different. The diagram to the right shows this condition. The diagram now reveals a disordinal interaction. So, an inquiring mind might wonder, is the interaction actually ordinal or disordinal? The answer is that if the lines cross under any circumstance, the interaction is disordinal, even though the disordinal interaction may have been "disguised" in one drawing. Though under most circumstances, disordinal interactions do not hide themselves, wise researchers should draw diagrams in more than one fashion before making final interpretations of interactions.



X_2 \ X_1	–	+
–	9	7
+	5	7

The reason identifying interactions is vital lies in the fact that the interpretations of results are completely different. When there is a disordinal or crossed interaction,

- researchers are not permitted to interpret the main effects involved in the interaction; and
- researchers acting on such information must look at the settings of two variables at once, rather than independent variables one at a time.

The reason for these rules may be obvious. With a disordinal interaction, each main effect from independent variables would accurately predict output about half the time. The rest of the time, results would not be as initially predicted from the main effects.[4] In short, interpreting the main effects would be misleading when there are disordinal interactions. Yet, this fact does not mean that the impact of the main effect is dropped from the study. The contribution to the total percentage of variance explained is retained. By the way, this requirement to avoid interpreting main effects for variables involved in disordinal interactions does not apply to *other* independent variable main effects that are *not* part of the disordinal interaction. If researchers had three independent variables, any main effect from the third variable could be interpreted provided it were not part of a disordinal interaction involving the first two variables.

If researchers wish to act upon their research findings, identifying the presence of disordinal interactions is very important. If there is an ordinal interaction, then it is possible to take action on one variable at a time, because the interaction is in the same direction as the main effects. On the other hand, if there is a disordinal interaction, researchers know that they cannot look at one independent variable at a time. If action is to be taken, one variable cannot be left to operate at random while the other is fixed.

COMPUTING THE FIXED-EFFECT ANOVA

The fixed-effect analysis of variance is "a model where the levels of the factor under study (the treatments) are *fixed* in advance. Inference is valid only for the levels under study" (Aczel, 1989, p. 379). In this approach, sometimes called the "Type I model" of analysis of variance, a serious choice is made of the levels of the independent variables, and the conditions are generally considered exhaustive of the variable. For instance, researchers might examine whether lengthy or brief speeches delivered by male or female speakers differ in their comprehensibility. Both the factors of speech length and sex of the speakers are composed of fixed levels that cover the researcher's range of interest. Most communication studies using factorial analysis of variance employ the fixed-effect model.

The fixed-effect model ANOVA makes four major assumptions that must be checked:

- Measurement of the dependent variable on the interval or ratio level
- Randomization
- Normal distribution of the dependent variable. Researchers may explore this matter by examining the kurtosis and skewness of the distributions. As shown in Chapter 4, researchers also could consider using the Kolmogorov-Smirnov or Shapiro-Wilks statistics to test the null hypothesis that the sample data do not differ from a normal

[4]For instance, in the previous example, there is a main effect from variable X_2 (though not from variable X_1). The main effect indicates that the dependent variable scores are highest when X_2 is set at its low level, but this effect will occur only if variable X_1 is low. If variable X_1 is high, then the dependent variable effects will be reduced when X_2 is at its high level.

distribution. With large sample sizes, the factorial analysis of variance is relatively robust to modest violations of normality. But if sample sizes are small (much below 30), researchers may wish to transform variables before continuing with analyses.

- Homogeneous variances. Researchers examine variances to detect if the variances are within the limits of sampling error and may use such tools as the F test (or F_{max}) of variances or Levene's test. If variances are unequal, researchers may look for ceiling or floor effects (by correlating means and variances). If heterogeneous variances exist in the absence of ceiling and floor effects, researchers have evidence of events by treatments interaction, and they should continue the search for moderating variables to include in the research effort.

The first two of these assumptions are matters that must be considered by researchers in their design or research. The third and fourth assumptions may be considered by examining data.

As mentioned in Chapter 8, the fixed-effect model also assumes these elements that are dependent on the research design choices and presumptions made by researchers:

- The elements in the model reflect the sum of all the elements that affect the dependent variable.
- The experiment contains all treatment levels of interest.
- The error effects are independent and normally distributed.
- The samples are independent, which means that knowledge of an individual's score on some measure neither predicts the degree of that individual's error nor affects the probability of predicting any other individual's responses.

Assessing Effects

Main effects for factorial analysis of variance are rather easy to compute. Researchers start by looking at the "between-groups variances" for each input variable separately. Table 9.2 shows such an example. To compute main effects mean squares is a relatively simple task because it uses the same formula as one-way analysis of variance, $n * s^2_{\bar{X}}$. Degrees of freedom for this term are the number of groups minus one. The process is repeated for the other main effects as well.

To compute interactions from levels of independent variables, researchers take each cell value and then subtract away the mean that would be expected by the row and column means (and means of "slices," in the case of three-way ANOVAs). Then, the contribution is contrasted by adding the grand mean (only once in the case of two-factor analysis of variance). Any remaining variation is attributable to interaction effects beyond influences of main effects alone. The degrees of freedom for interaction effects are simply the product of the degrees of freedom for each main effect involved in the interaction.

Within-groups variance (also known as error variance) is the same as a pooled variance s_p^2 introduced in Chapter 8. If sample sizes are equal, the pooled variance is simply the mean of the variances within each cell group. If the sample sizes are unequal, a formula that adjusts for sample sizes is used.

(Text continues on page 225)

Table 9.2 Example of a Two-Way Analysis of Variance

A researcher was interested in student's perceptions of their teachers' overall verbal communication and nonverbal immediacy with students (measured on a set of 18 seven-point "communication satisfaction" scales, with a possible range for the total of 18 to 126). The researcher thought that female students might report greater communication of teacher immediacy than male students generally. The researcher also suspected that students who were highly individualistic by nature (measured by a standard scale and then divided into "high" and "low" groups based on their scores) might report greater communication of teacher immediacy than would students who were low on individualism. The hypothesis for the sex of student was H_1: $\mu_{\text{women}} > \mu_{\text{men}}$. The hypothesis about individualism predicted H_2: $\mu_{\text{highly individualistic students}} > \mu_{\text{lowly individualistic students}}$. The researcher had a number of male and female students complete scales to rate their individualism and their perceptions of teacher verbal and nonverbal communication of immediacy. Their scores are found in the grid below, along with a list of means ($\bar{X}$) and variances (s^2). To compute the sources of variation, researchers must examine two main effects and an interaction effect.

		Sex of Student		
		Male	*Female*	
Individualism	*Low*	102, 103, 95, 80, 95, 110, 109, 102, 94, 98, 90, 101, 108, 105, 90, 96, 90, 93, 111, 95, 88, 81, 87, 91, 85, 89 $\bar{X}_{11} = 95.69$ $s^2_{11} = 76.30$	97, 86, 104, 95, 92, 82, 91, 99, 74, 106, 90, 109, 89, 104, 101, 102, 71, 86, 89, 112, 90, 94, 94, 86, 107, 94 $\bar{X}_{12} = 94.00$ $s^2_{12} = 102.80$	$\bar{\bar{X}}_{.1} = 94.85$
	High	114, 105, 104, 110, 76, 87, 84, 105, 85, 96, 91, 109, 95, 98, 101, 83, 99, 75, 103, 113, 101, 93, 102, 116, 104, 103 $\bar{X}_{21} = 98.15$ $s^2_{21} = 125.42$	76, 101, 112, 104, 100, 109, 104, 99, 100, 116, 108, 106, 90, 111, 90, 110, 105, 102, 103, 92, 98, 114, 85, 104, 89, 94 $\bar{X}_{22} = 100.85$ $s^2_{22} = 91.74$	$\bar{\bar{X}}_{.2} = 99.50$
		$\bar{\bar{X}}_{.1} = 96.92$	$\bar{\bar{X}}_{.2} = 97.43$	$\bar{\bar{\bar{X}}} = 97.18$

- The mean square for individualism is computed as $MS_{\text{individuals}} = n$ in the row $* s^2_{\overline{X}}$ or, equivalently using the computational formula with *j* as the rows (individualism), *k* as the columns (sex of subjects), and *i* as instances of scores:

$$\left(\frac{\sum_j \left(\sum_k \sum_i X_{ijk} \right)^2}{n_{j.}} - \frac{\left(\sum_k \sum_j \sum_i X_{ijk} \right)^2}{n_{..}} \right) \Big/ \text{degrees of freedom.}$$

Using the conceptual formula, one may substitute scores as $MS_{\text{individuals}} = n$ in the row $* s^2_{\overline{X}}$. The equation then can be shown:

$$\begin{aligned}
&= 52 * \frac{\sum \left(\overline{\overline{X}}_j - \overline{\overline{\overline{X}}} \right)^2}{j - 1} \\
&= 52 * \frac{(94.85 - 97.18)^2 + (99.5 - 97.18)^2}{2 - 1} \\
&= 52 * \frac{5.43 + 5.38}{1} \\
&= 52 * 10.81 \\
&= 562.12
\end{aligned}$$

- The mean square for sex of participants are computed as $MS_{\text{sex of participants}} = n$ in the column $* s^2_{\overline{X}}$ or, equivalently using a computational formula with *j* as the rows (individualism), *k* as the columns (sex of participants), and *i* as instances of scores:

$$\left(\frac{\sum_k \left(\sum_j \sum_i X_{ijk} \right)^2}{n_{.k}} - \frac{\left(\sum_k \sum_j \sum_i X_{ijk} \right)^2}{n_{..}} \right) \Big/ \text{degrees of freedom.}$$

Using the first formula, one may substitute scores as $MS_{\text{sex of participants}} = n$ in the column $s^2_{\overline{X}}$ to get the following equality:

$$\begin{aligned}
&= 52 * \frac{\sum \left(\overline{\overline{X}}_k - \overline{\overline{\overline{X}}} \right)^2}{k - 1} \\
&= 52 * \frac{(96.92 - 97.18)^2 + (97.43 - 97.18)^2}{2 - 1} \\
&= 52 * \frac{.07 + .06}{1} \\
&= 52 * .13 \\
&= 6.76.
\end{aligned}$$

(Continued)

Table 9.2 (Continued)

- The mean square for the interaction between individualism and sex of subjects in this case is computed as $MS_{interaction}$ = (n in each cell condition $* \sum$ [cell mean – row mean – column mean + grand mean]^2) / degrees of freedom or, equivalently using the computational formula with j as the rows (individualism) k as the columns (sex of subjects), and $\overline{\overline{\overline{X}}}$ as the grand mean:

$$\left(n_{jk} * \sum_j \sum_k \left(\overline{X}_{jk} - \overline{\overline{X}}_{j.} - \overline{\overline{X}}_{.k} + \overline{\overline{\overline{X}}}\right)^2\right) \text{/degrees of freedom.}$$

  Substituting the scores, one finds:

$$= \left(26 * \begin{bmatrix} (95.69 - 94.85 - 96.92 + 97.18)^2 + (94 - 94.85 - 97.43 + 97.18)^2 + \\ (98.15 - 99.5 - 96.92 + 97.18)^2 + (100.85 - 99.5 - 97.43 + 97.18)^2 \end{bmatrix}\right) /1$$
$$= (26 * [1.21 + 1.21 + 1.19 + 1.21])/1 = 125.32.$$

- Within groups variance (also known as error variance) is the same as a pooled variance s_p^2. If sample sizes are equal, the pooled variance is simply the mean of the variances within each cell group. If the sample sizes are unequal, the following formula that adjusts for sample sizes is used

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots (n - 1)s_n^2}{n_1 - 1 + n_2 - 1 + \ldots n_n - 1}.$$

  For these data, the pooled variance is 99.07. This term is the mean square, not the sums of squares. When entered into the ANOVA table, the numbers are revealed as:

Sources of Variation	*Sums of Squares*	*Degrees of Freedom*	*Mean Square*	*F*	*Eta Squared*
Individuality	562.12	1	562.12	5.67*	.05
Sex of participants	6.76	1	6.76	<1	
Individuality x sex of subjects	125.32	1	125.32	1.26	
Within-groups variance	9907	100	99.07		
Within-groups variance	9907	100	99.07		
Total	1,0601.2	104			

*$p < .05$.

- The degrees of freedom for each main effect are the number of levels of the variable, minus one. The degrees of freedom for the interaction effect are determined by multiplying the main effect degrees of freedom for each variable involved in the interaction. Hence, in this case, degrees of freedom for the interaction term equal one (1 * 1 = 1). The within-groups degrees of freedom are the total number of events minus the number of groups in the study. In this case, there were 104 events, minus four sample groups of cells. Thus, degrees of freedom equal 100.
- To test the statistical hypothesis of "no difference" with an alpha risk of .05, the researcher finds the critical value of *F* with the numerator degrees of freedom from the effect and degrees of freedom in the denominator equal to the number of degrees of freedom for the within-groups (or "error") term. In this case, with 1 and 100 degrees of freedom, the critical *F* ratio is 3.936. Thus, the main effect from individualism was statistically significant.
- To identify the source of differences from the main effect, one need only look at the marginal means (because there are only two levels). Those with high individualism showed the highest ratings communication immediacy. This effect was associated with 5% of the total variance.
- In this case, there was no statistically significant interaction effect. Yet, if a statistically significant interaction had been found, the researcher would want to examine diagrams of the specific interaction patterns.
- To investigate whether the assumptions underlying the analysis of variance held, the homogeneity of variance was tested. In this case, the *F* test was

$$F = \frac{\text{largest } s^2}{\text{smallest } s^2} = \frac{102.90}{76.30} = 1.35.$$

The critical $F_{(d.f.:\ 25,25)}$ with α at .025 was 2.23. Thus, the assumption of homogeneous variance held.

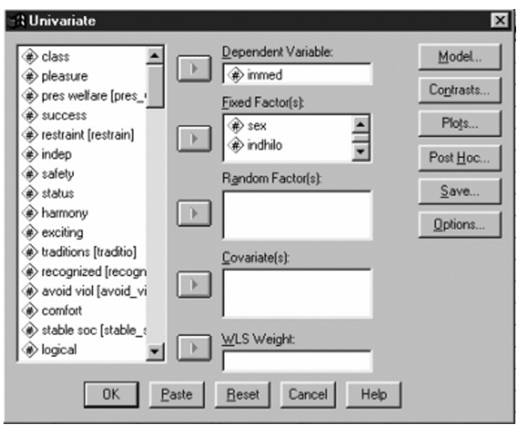
After the factorial analysis of variance is completed, researchers would be well advised to determine the size of each significant effect. The formula for eta squared (η^2) used for the one-way analysis of variance may be used here. The formula is

$$\eta^2 = \frac{\text{between-groups sums of squares}}{\text{total sums of squares}}.$$

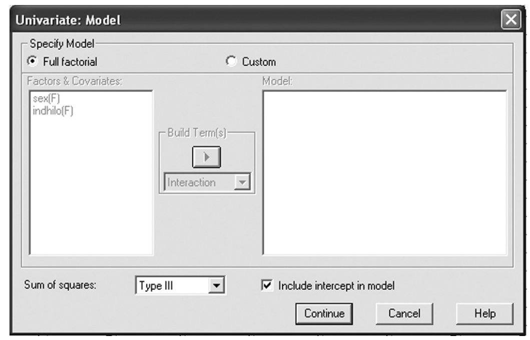
In each case, the "between-groups sums of squares" are the sums of squares associated with each individual statistically significant effect, whether a main or interaction effect. This convenient formula permits the researcher to determine the share of total variance that is contributed by each individual source that produces a significant difference.

Using SPSS for Fixed-Effect ANOVA

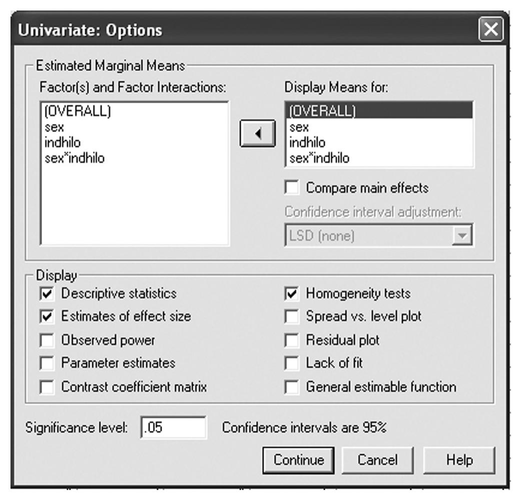
As an illustration of a fixed-effect factorial analysis of variance, data from a study of perceived teacher immediacy will be considered. From the *Analyze* menu, the researcher would choose *General Linear Model* and then *Univariate. . . .* The dependent variable "immediacy" would be identified and moved to the "Dependent Variable:" field. The two independent variables, "sex" and "indhilo" (individualism) of the respondents, are highlighted and transferred to the "Fixed Factor(s):" field.



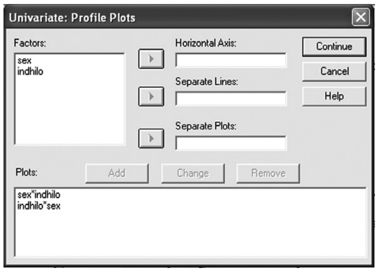
The researcher then would click on the *Model. . .* button to specify the "Full factorial" option (if it has not already been identified). The dialog box also includes the method to be used to compute sums of squares. The default is Type III, which includes a full factorial with no missing data cells. This reference to Type III refers only to the sums of squares, *not* the type of factorial design. Afterward, the researcher would click on the *Continue* button. (See illustration on the top right column.)



The researcher then clicks on the *Options* to identify the effects to be tested. In this case, the researcher requests displays of means for all effects by transferring the effects from the "Factor(s) and Factor Interactions:" field to the "Display Means for:" field. In addition, boxes are checked to obtain "Descriptive statistics," "Estimates of effect size," and "Homogeneity [of variance] tests." Afterward, the researcher clicks the *Continue* button.



To obtain a plot of the results, the researcher would click on the *Plots. . .* button. Because it is wise to include two diagrams with different variables placed on the horizontal axis, two plots are requested. The first plot, "sex*indhilo," places "sex" on the horizontal axis. The second plot, "indhilo*sex," places "indhilo" on the horizontal axis. The researcher would click on *Continue* and then *OK.* (See illustration on the top of left column on p. 227.)



Levene's test was not statistically significant. Hence, the researcher would conclude that the assumption that the variances were the same could not be rejected.

Univariate Analysis of Variance

Levene's Test of Equality of error Variances

Dependent Variable: IMMED

F	df1	df2	Sig.
.503	3	100	.881

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

a. Design: Intercept+ SEX+INDHILO+SEX*INDHILO

These steps would produce the following results:

Descriptive Statistics

Dependent Variable: IMMED

SEX	INDHILO	Mean	Std. Deviation	N
1	1	95.69	8.74	26
	2	98.15	11.20	26
	Total	96.92	10.02	52
2	1	94.00	10.14	26
	2	100.85	9.58	26
	Total	97.42	10.36	52
Total	1	94.85	9.41	52
	2	99.50	10.41	52
	Total	97.17	10.15	104

Tests of Between-Subjects Effects

Dependent Variable: IMMED

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Mode	694.577[a]	3	231.526	2.337	.078
Intercept	962031.115	1	982031.115	9913.190	.000
SEX	6.500	1	6.500	.066	.798
INDHILO	563.115	1	563.115	5.684	.019
SEX*INDHILO	124.962	1	124.962	1.261	.264
Error	9906.308	100	99.063		
Total	992632.000	104			
Corrected Total	10600.885	103			

a. R Squared = .066 (Adjusted R Squared = .037)

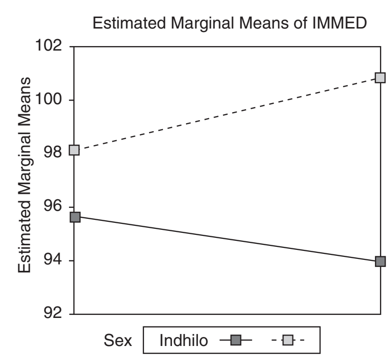
The ANOVA table indicated that the individualism variables produced a significant difference and that no significant interaction effect was present. In the estimated marginal means, those with high mean ratings of individualism (Group 2) reported greater perceptions of the teacher immediacy than others (Group 1).

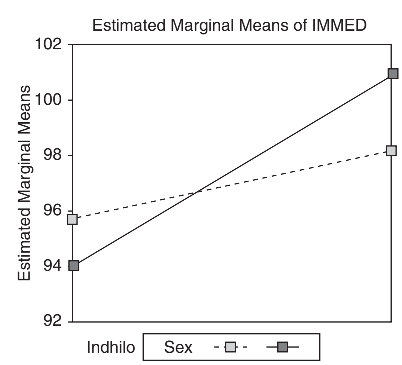
3. INDHILO

Dependent Variable: IMMED

INDHILO	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	94.846	1.380	92.108	97.585
2	99.500	1.380	96.762	102.238

Although the ANOVA showed that that interactions were not statistically significant, it is interesting to note that the type of interaction would have been disordinal because when the chart was drawn with individualism on the horizontal axis, the result was crossed interaction, as shown in the second chart below.



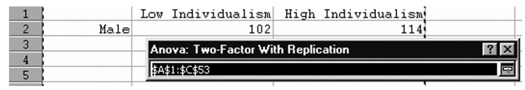


Using Excel for Fixed-Effect ANOVA

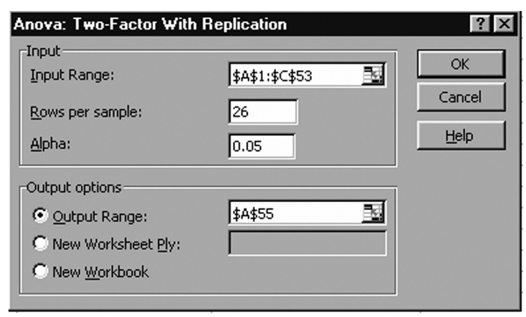
When completing a "two-way" or two-factor analysis of variance, Excel requires that each level of one independent variable should be identified by separate columns. The rows must group the levels of the dependent variable. So, the data must be arranged in a way similar to the one shown below. In addition, the rows must group the scores of the dependent variable.

	A	B	C
1		Low Individualism	High Individualism
2	Male	102	114
3		103	105
4		95	104
5		80	110
6		97	76
7		110	87
8		109	84
9		102	105
10		94	85
11		98	96
12		90	91
13		101	109
14		108	95
15		105	98
16		90	83
17		96	99
18		90	75
19		93	103
20		111	113
21		95	101
22		88	101
23		81	93
24		87	102
25		91	116
26		85	104
27		89	103
28	Female	97	76
29		86	101
30		104	112

To complete the two-way ANOVA, the researcher selects *Data Analysis* from the *Tools* menu. From the dialog box that appears, the researcher selects "Anova: Two-Factor With Replication" and then clicks the *OK* button. In the box that appears, the researcher clicks on and then highlights the location on the spreadsheet where the data begin through the location where the data end.



The researcher clicks on to return to the dialog box. To select a location to place output, the researcher clicks on the "Output Range:" window and goes to an available location such as one starting at cell A55. After making a selection, the researcher clicks on to return to the dialog box. In the dialog box, the researcher must specify a desired alpha risk for use in significance testing (.05 in this case). In addition, the researcher identifies the "Rows per sample:" a number that corresponds to the sample size in each cell.



After clicking on the *OK* button, the following output appears. Inspecting the column marked *P-value* makes it clear that the only significant effect was a main effect for the column variable, individualism. A look at the means of the high and low individualism groups reveals that the highest ratings (dependent variable: immediacy) were from the high individualism group participants. Excel does not have built-in functions for designs involving more than two independent variables, but this chapter's Web site includes references to other add-in programs that help fill these additional needs.

Anova: Two-Factor With Replication						
Summary	Low Individualism	High Individualism	Total			
Male						
Count	26	26	52			
Sum	2488	2552	5040			
Average	95.69230769	98.15384615	96.92307692			
Variance	76.30153846	125.4153846	100.4253394			
Female						
Count	26	26	52			
Sum	2444	2622	5066			
Average	94	100.8461538	97.42307692			
Variance	102.8	91.73536462	107.3076923			
Total						
Count	52	52				
Sum	4932	5174				
Average	94.84615385	99.5				
Variance	88.52488689	108.2941176				
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	6.5	1	6.5	0.06561476	0.798358302	3.936150961
Columns	563.1153846	1	563.1153846	5.684412418	0.019002603	3.936150961
Interaction	124.9615365	1	124.9615385	1.261434051	0.264069015	3.936150961
Within	9906.307692	100	99.06307692			
Total	10600.88462	103				

RANDOM- AND MIXED-EFFECTS DESIGNS

Though in "fixed-effects" analysis of variance designs, the levels of a variable are *fixed* to include the full range of levels that a researcher considers important, there are times when researchers take random draws of examples from populations as variable levels. There also are times when the same participants are measured repeatedly.[5] This discussion will consider each of these approaches.

Understanding Random-Effects Designs

In analysis of variance, the **random effects model** is an experimental design in which the levels of the factor are random, in the sense that they are drawn at random from a population of levels rather than fixed by an investigator. Also called "variance components model" and "Model II ANOVA design,"

> [t]he random-effects model is used when there is a large number of categories or levels of a factor. For example, say researchers in a survey organization wanted to see whether different kinds of telephone interviewers get different response rates. Because there are [*sic*] potentially a very large number of categories (difference in accent, quality of voice,

[5]This is not meant to equate random effects with repeated measures. Whereas random-effects testing involves further partialing of the within-subjects error term, random effects do not require repeated measures designs.

> etc.), perhaps as many as there are individual telephone interviewers, a sample is chosen randomly from the population of interviewers, which is also, in this case, a population of levels. On the other hand, if the survey organization were interested only in, say, the difference in response rate between male and female interviewers, they would use a fixed-effects model. (Vogt, 2005, pp. 260–261)

In this model, the researcher attempts to make an inference about the meaningfulness of an entire population of a variable. If a researcher finds that that variable fails to produce statistically significant effects, the researcher may have some reason to move to other variables that actually might produce different effects in the dependent variable. For most research in communication studies, the random-effects model is not used; instead, either a fixed- or mixed-effects model is employed.

For the random-effects model, several assumptions must be made in addition to the ones typically required of any parametric tests (i.e., interval or ratio level measurement of the dependent variable, randomization, normal distribution, homogeneity of variances). In particular, this model assumes that the error terms are independent and normally distributed. This assumption means, among other things, that knowing the factor level would not allow one to predict whether an error term was above or below zero.

Computationally, a one-factor fixed-effects model and the one-factor random-effects model are equivalent, but they differ in the additional assumption (independent and normally distributed error terms) and the interpretation of the meaning of an effect. The existence of a statistically significant effect means that the variable from which random selections of levels were made produces significant effects on the dependent variable. Hence, the variable bears further inquiry to determine fixed levels that produce predictable outcomes. It is, therefore, a variable "screening" exercise. It should be added that the computation of statistical significance for the random-effect variable does change when two-factor mixed-effect designs are involved. These highly useful designs are considered next.

Use of Mixed-Effects Designs

Sometimes researchers employ one or more fixed effects and one or more random effects. The mixed-effects ANOVA (also known as "Model III ANOVA") combines (mixes) between-subjects factors and within-subjects factors.

The Logic of Combining Fixed and "Random" Effects

One might wonder why communication researchers would want to complete studies where some variables are fixed and others are treated as randomly selected examples of variable levels. For example, suppose a researcher presented participants with four messages with or without emotional language (the fixed effect). Furthermore, suppose the researcher used messages on two different topics. If the researcher were interested in these specific messages, the message topics would be a fixed effect. If instead the researcher were interested in the influences of messages in general, the message topics would be a random effect. Statistically, if one uses a completely fixed-effect ANOVA, the "between-groups variance term" includes both the treatment and reactions to the specific message "replications" used. Some scholars have recommended that researchers who study message characteristics should include at least two examples of message and more than one topic (Clark, 1973; Jackson, Brashers, & Massey, 1992; Jackson & Jacobs, 1983; Jackson, O'Keefe, & Jacobs, 1988; Jackson, O'Keefe, Jacobs, & Brashers, 1989).[6] A remaining controversy is whether the effect should be treated as a random selection of a variable level from the population of theoretic possibilities, or whether the effect should be considered a fixed effect and studied as any other primary variable of interest. If a rationale can justify the use of a fixed-effect approach, that method probably should be used. But if the researcher is using multiple messages simply for the purpose of control, the mixed-effect ANOVA is a particularly appropriate tool.

In the mixed-effects model, the fixed effect and interaction mean squares are identified the same way they are for the fixed-effect model. But for the computation of the *F* statistic, things change. In the mixed-effects model, each main effect mean square is divided by the *interaction* mean square between the fixed-effect variable and the random-effect variable. The use of the interaction mean square allows the shared contribution of the random influence to be controlled in estimates of the primary fixed effects of interest. For the interaction term itself, the within-groups mean square is used to compute the *F* value. Naturally, the degrees of freedom are computed differently depending on whether the within-groups mean square or the interaction mean square is used as the divisor:

- For the fixed effect, the degrees of freedom are

$$\frac{\text{[for the numerator] number of levels in the fixed effect} - 1}{\begin{array}{c}\text{[for the denominator] degrees of freedom in the interaction term}\\ \text{(number of levels in the fixed effect} - 1) * \text{(number of levels}\\ \text{in the random variable} - 1)\end{array}}.$$

- For the random effect, the degrees of freedom are

$$\frac{\text{[for the numerator] number of levels in the random effect} - 1}{\begin{array}{c}\text{[for the denominator] degrees of freedom in the interaction term}\\ \text{(number of levels in the fixed effect} - 1) * \text{(number of levels}\\ \text{in the random variable} - 1)\end{array}}.$$

- For the interaction effect, the degrees of freedom are

$$\frac{\begin{array}{c}\text{[for the numerator] (number of levels in the fixed effect} - 1) *\\ \text{(number of levels in the random variable} - 1)\end{array}}{\begin{array}{c}\text{[for the denominator] degrees of freedom in the within-groups}\\ \text{term (number of events in the study} - \text{number of groups)}\end{array}}.$$

---

[6] At one time, communication researchers routinely studied message variables by presenting participants with one example of a message and a message topic. Critics suggested that it often was difficult to know if the results could be generalized beyond the specific message and the specific topics. This position is controversial. Though admitting the value of replications, some have disagreed that using single messages is a flawed approach (especially when a series of studies across many topics has been completed) (Burgoon, Hall, & Pfau, 1991). Furthermore, some argue that insisting on multiple message replications may be an unfair demand that may drive out most worthy message-effects research. Settling the matter is not within the purview of this book. Yet, this chapter accepts the view that all researchers must shoulder a burden of proof and present arguments and reasons to support the proposition that the message examples they study are representative of some meaningful class of messages.

As can be seen, if there is a great interaction between the fixed effect and the random effect, the chances of finding a statistically significant fixed effect are reduced. But if the interaction is modest, the test can be quite powerful. Table 9.3 shows the steps involved in such a mixed-effects study.

Table 9.3 Example of Mixed-Effects Model

A researcher wanted to know if a message using deliberate ambiguity would produce more attitude change among initially hostile audiences than messages with clear language. The hypothesis was

H_1: $\mu_{ambiguous} > \mu_{unambiguous}$.

But the researcher wanted to determine if the strategies, rather than the particular topic of the messages, were responsible. So, the use of the message strategies was the fixed effect, and three different messages were designed with and without the ambiguous language used.

	Message Topic 1	*Message Topic 2*	*Message Topic 3*	$\bar{X}_{j.}$	$\bar{X}_{j.} - \bar{\bar{X}}$
Unambiguous language	29, 20, 22, 33	36, 23, 25, 24	38, 26, 21, 27	$n = 12$	
$\bar{X}_{1k}$	26	27	28	27	−3.165
s^2_{1k}	36.67	36.67	51.33		
Ambiguous language	42, 32, 25, 29	31, 45, 27, 37	31, 31, 25, 45	$n = 12$	
$\bar{X}_{2k}$	32	35	33	33.33	3.165
s^2_{2k}	52.67	61.33	72		
$\bar{X}_{\cdot k}$	$n = 8$	$n = 8$	$n = 8$		
	29	31	30.5	30.165	
$\bar{X}_{\cdot k} - \bar{\bar{X}}$	–1.165	.835	.335		

These messages were approximately the same length and had human interest quotients of nearly the identical levels. Computing sources of variation involves both fixed and random effects:

$$\text{Mean Square}_{\text{fixed effect}} = \frac{n_j * \sum\left(\bar{X}_j - \bar{\bar{X}}\right)^2}{j - 1},$$

where

n_j is the number of events in each level of the fixed effect and

j is the number of levels for the fixed effect.

In this case, the numbers are

$$\text{Mean Square}_{\text{fixed effect}} = \frac{12 * (-3.165^2 + 3.165^2)}{2-1} = \frac{12 * 20.03}{1} = 240.41$$

$$\text{Mean Square}_{\text{random effect (messages)}} = \frac{n_k * \sum \left(\overline{X}_k - \overline{\overline{X}}\right)^2}{k-1},$$

where

n_k is the number of events in level of the random effect and

k is the number of levels for the random effect.

In this case, the data entry yields

$$\text{Mean Square}_{\text{random effect (messages)}} = \frac{8 * (-1.165^2 + .885^2 + .335^2)}{3-1} = \frac{8 * 2.165}{2} = 8.66$$

$$\text{Mean Square}_{\text{interaction effect (messages)}} = \frac{n_{jk} * \sum \left(\overline{X}_{jk} - \overline{X}_{j.} - \overline{X}_{.k} + \overline{\overline{X}}\right)^2}{(j-1) * (k-1)},$$

where

n_{jk} is the number of events in each cell,

j is the number of levels for the fixed effect,

k is the number of levels for the random effect,

$\overline{X}_{jk}$ is the mean of each individual cell,

$\overline{X}_{j.}$ is the mean of the corresponding level of the fixed effect,

$\overline{X}_{.k}$ is the mean of the corresponding level of the random effect, and

$\overline{\overline{X}}$ is the grand mean.

In this case, the data entry yields

$$\text{Mean Square}_{\text{interaction effect (messages)}} = \frac{4 * [(26.67 - 27 - 29 + 30.165)^2 + \ldots (33 - 33.33 - 30.5 + 30.165)^2]}{(2-1) * (3-1)}$$

$$= \frac{4 * 2.33}{2} = \frac{9.32}{2} = 4.66$$

$$\text{Mean Square}_{\text{within groups}} = \frac{\sum s_{jk}^2}{j * k},$$

(Continued)

Table 9.3 (Continued)

where

s_{jk}^2 is variance within each cell,

j is the number of levels for the fixed effect, and

k is the number of levels for the random effect.

In this case, the data are

$$\text{Mean Square}_{\text{within groups}} = \frac{36.67 + 36.67 + 51.33 + 52.67 + 61.33 + 72}{2 * 3} = \frac{310.67}{6} = 51.78.$$

The ANOVA table for this effect is

Source	*SS*	*d.f.*	*MS*	*F*
Ambiguity (fixed)	240.41	1	240.41	51.59
Messages (random)	17.32	2	8.66	1.86
Interaction	9.32	2	4.66	.09
Within groups	9.32	18	51.78	

To compute the *F* ratios for both the fixed and random effects, the mean squares are divided by the interaction mean square. Therefore, the degrees of freedom to identify critical *F* values are for the fixed and random effects in the numerator and the interaction term for the denominator. Thus, the degrees of freedom for ambiguity are 1 in the numerator and 2 in the denominator. With alpha risk of .05, the critical *F* ratio is 18.5.

For the random effect, degrees of freedom are 2 for the numerator and 2 for the denominator. The critical *F* ratio is 19 with alpha risk at .05.

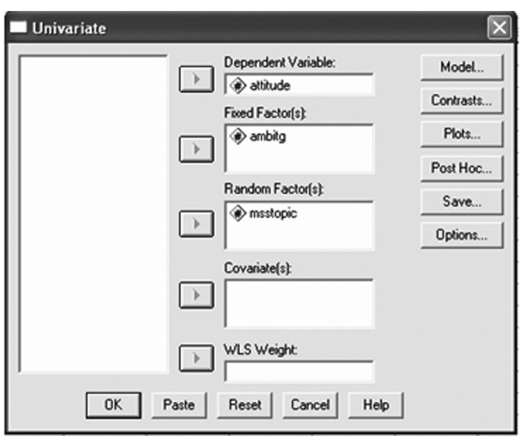
For the interaction term, the within-groups mean square is used in the denominator to compute the *F* ratio. Thus, the degrees of freedom are 2 and 18. This critical *F* ratio is 3.55.

In addition to the assumption of parametric statistics generally (interval or ratio level measurement of the dependent variable, normal distribution, randomization, and homogeneity of variance), the mixed effects ANOVA also assumes the following:

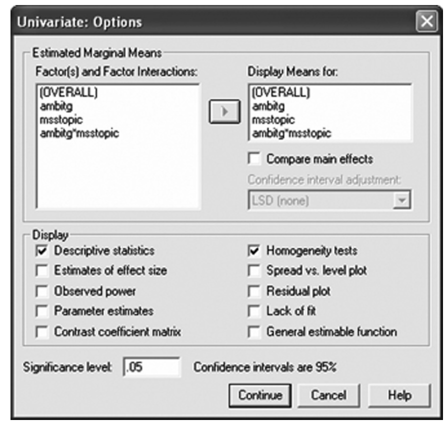
- Random effects, error effects, and interaction effects are normally distributed.
- Compound symmetry is made. This assumption holds that the covariance matrices for the levels of the "between factor" are homogeneous. Though a test for this assumption is available (Box, 1950), trusting routine testing for this assumption is not widely recommended (Kirk, 1982, p. 503). In addition, though violating this assumption slightly increases true Type I error, the effects tend to be small (Collier, Baker, Mandeville, & Hayes, 1967).

Using SPSS for Mixed-Effects ANOVA

In an example of a mixed model, Table 9.3 shows a study involving a researcher who examined whether use of deliberately ambiguous language in two sample speeches presented to initially hostile audience members would produce more positive attitudes than would the use of clear statements. The three messages would be random factors, and the strategic message ambiguity would be the fixed factor. To complete such analysis in SPSS, a researcher would select *General Linear Model* from the *Analyze* menu followed by *Univariate.* In the dialog box that appears, the researcher would transfer the appropriate dependent, independent, and random factors. The researcher clicks on the *Model. . .* button to specify the "Full Factorial" option (if it has not already been identified). Then the researcher clicks on the *Continue* button.



In the *Univariate* dialog box, the researcher clicks on the *Options. . .* button. The researcher highlights all the effects and transfers them to the "Display Means for:" field. To complete the report, the "Descriptive statistics" and "Homogeneity tests" boxes should be checked. The researcher then clicks on the *Continue* and *OK* buttons.



The results appear in the output window. As can be seen by the results below, the test of homogeneous variances was not statistically significant. Thus, the assumption of homogeneous variances was considered tenable.

Descriptive Statistics

Dependent Variable: ATTITUDE

AMBITG	MSSTOPIC	Mean	Std. Deviation	N
unambig	1	26.00	6.06	4
	2	27.00	6.06	4
	3	28.00	7.16	4
	Total	27.00	5.89	12
ambig	1	32.00	7.26	4
	2	35.00	7.83	4
	3	33.00	8.49	4
	Total	33.33	7.24	12
Total	1	29.00	6.97	8
	2	31.00	7.76	8
	3	30.50	7.75	8
	Total	30.17	7.22	24

Levene's Test of Equality of Error Variance's

Dependent Variable: ATTITUDE

F	df1	df2	Sig.
.103	5	18	.990

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

a. Design: Intercept+AMBITG+MSSTOPIC+AMBITG *MSSTOPIC

The ANOVA table revealed a statistically significant main effect from the deliberate ambiguity fixed effect. The previously presented means showed that the unambiguous messages produced less favorable attitudes toward the topic than the deliberately ambiguous messages. As the researcher hoped, the random effect "msstopic" was not statistically significant, indicating that the effects were related to the message strategy across the selection of message examples.

Tests of Between-Subjects Effects

Dependent Variable: ATTITUDE

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	21840.667	1	21840.667	2520.077	.000
	Error	17.333	2	8.667[a]		
AMBITG	Hypothesis	240.667	1	240.667	51.571	.019
	Error	9.333	2	4.667[b]		
MSSTOPIC	Hypothesis	17.333	2	8.667	1.857	.350
	Error	9.333	2	4.667[b]		
AMBITG * MSSTOPIC	Hypothesis	9.333	2	4.667	.090	.914
	Error	932.000	18	51.778[c]		

a. MS(MSSTOPIC)

b. MS(AMBITG * MSSTOPIC)

c. MS(Error)

The Repeated Measures Design

An interesting application of the mixed model design is called the **repeated measure ANOVA**, in which each experimental unit (person or item) is assigned to all treatments of at least one fixed factor. Then, several different observations are made of them. For instance, researchers could take communication apprehension measures from students in a public speaking class several times during the semester. Then, the repeated measures could be referenced. Similarly, researchers could expose people to three or more message treatments (with and without evidence, for instance) in random order as part of a counterbalanced design. The order in which the same people received the message could be tracked as a repeated measure. If there were any significant fatigue effects, they might be studied directly. In short, communication scholars may use this method in very creative ways.

For conceptual purposes, it makes sense to imagine a simple case. If people were surveyed at different times, shifts in their measured scores could be explained over time. Statistically significant differences could be tracked as repeated measures. The repeated measure is, in fact, treated as a fixed effect. The participants themselves would be used as a random effect because they would be selections of people from the larger population. If the fixed effect were diagrammed in a column and the random effect (the different participants) assigned to individual rows, there would be only one participant in each of the cells. It would not really be possible to determine a within-groups variance. Thus, just as with a mixed-effect design, the researcher could enlist the interaction term to substitute for the within-groups variance. Special Discussion 9.2 shows an example of this repeated measure design.

Special Discussion 9.2

Counterbalanced Designs

There are many times that participants in a study must be exposed to a series of treatments. They may be presented with three or more messages with different persuasive appeals or language treatments. These exposures could stimulate change in subjects even though the fixed effects may not be responsible. There are several ways that it may occur:

- Subjects may become “testwise” by the end of the study rather than remaining relatively “naive” as they were at the beginning.
- Subjects may grow fatigued over time.
- Cumulative exposure to content may make a subject increasingly sensitive to message elements by the end of the study.

There are several ways to deal the problem, such as lengthening the time between exposures to treatments or reducing the numbers of treatments subjects receive. Another option involves counterbalancing, which presents “conditions (treatments) in all possible orders to avoid order effects” (Vogt, 2005, p. 67). In short, the researcher randomly assigns subjects to receive the treatments in random order. A variation of this form is called the Latin square design, which is “a method of allocating subjects, in a within-subjects experiment, to treatment group orders. So called because the treatments are symbolized by Latin (not Greek) letters” (Vogt, 2005, p. 169). For the Latin square, the number of rows and columns must be equal (a square). For instance, researchers might counterbalance four treatments (A, B, C, and D) in the following fashion.

	Order of Presentation			
Person 1	A	B	C	D
Person 2	B	C	D	A
Person 3	C	D	A	B
Person 4	D	A	B	C

This method ensures that the order effects are “balanced” so that order biases do not affect reactions to one treatment any more than any other. The influence of order does not disappear, but it is mixed (confounded) with measures of error variance.

Alternatively, the data may be analyzed (and usually will be) by use of a mixed-effects model ANOVA. This analysis of data can grow increasingly complex as the researcher interprets the meaning of the independent variable as a within-groups variable and may include the counterbalancing sequence as an added between-subjects factor. Thus, in a mixed-effects design, the number of effects to examine may increase.

The repeated measures ANOVA embodies several assumptions. In addition to interval or ratio level measurement of the dependent variable, normal distribution, randomization, and homogeneity of variances, two other assumptions are featured prominently. The first of these is the independence of observations. There is no way to test the matter. Instead, the researcher must use a sound experimental design with a different person for each "row" in the random effect.

A second assumption is sphericity. The notion is related to the requirement of compound symmetry in mixed-effect designs. One scholar explains:

> For many years it was thought that stronger condition, called uniformity (compound symmetry), was necessary. The uniformity condition required that the population variances for all treatments be equal and also that all population covariances are equal. However, Huynh and Feldt (1970) and Rouanet and Lepine (1970) showed that sphericity is an exact condition for the *F* test to be valid. Sphericity requires that the variances of the difference for *all* pairs of repeated measures be equal. (J. P. Stevens, 2002, p. 500)

Of course, if a matrix has compound symmetry, it will meet the sphericity assumption by definition.

Sphericity is the assumption "of independent observations with a constant variance" (Upton & Cook, 2002, p. 344). For instance, if a researcher took listening ability measures of interpersonal communication students for three consecutive weeks, there would be a difference between Week 1 and Week 2 and between Week 2 and Week 3. In essence, two new variables would be created for analysis. If these new variables were uncorrelated with each other, sums of squares would equal zero. The matrix of such data would be called **orthonormal**. When a matrix of the new variables and the covariance of the original variables are compared, then (using the language of matrix operations) "the sphericity assumption says that the covariance matrix for the new (transformed) variables is a diagonal matrix with equal variance in the diagonal. . . . Saying that the off diagonal elements are 0 means that the covariance for all transformed variables are 0, which implies that the correlations are 0" (J. P. Stevens, 2002, p. 501). If this assumed sphericity is not present, then the *F* test statistic tends to be inflated. In other words, researchers will mistakenly reject null hypotheses more often than their announced alpha risks.

To test for sphericity, Mauchly's test usually is involved. This test is given by

$$W = \det(\mathbf{S}) \left(\frac{k+1}{\mathrm{tr}(\mathbf{S})} \right)^{k+1},$$

where

$\mathbf{S}$ is a $k \times k$ sample covariance matrix,

$\det(\mathbf{S})$ is the determinant of the $k \times k$ covariance matrix, and

$\mathrm{tr}(\mathbf{S})$ is the trace of the $k \times k$ covariance matrix.

Because this formula is mildly complicated, no computation example will be shown here. Instead, this test will be revisited in the section on using SPSS to help analyze data. After the W is computed, the "chi-square" (χ^2) distribution may be used to assess whether the data show a statistically significant difference from sphericity. If the observed chi-square value is greater than the required critical value at the specific alpha risk set by the researcher, the null hypothesis of sphericity is rejected.

This measure reports if there is a correlation between pairs of repeated measures. If the observed statistic for Mauchly's test is statistically significant and the underlying distributions are normal, then the assumption of sphericity is rejected and an adjustment will be necessary. It may be useful to switch to the multivariate form of repeated measures ANOVA in which sphericity is not assumed. It should be mentioned that although Mauchly's test is a popular tool, it has been criticized for its inaccuracy when multivariate normality cannot be assured (Keselman, Rogan, Mendoza, & Breen, 1980; Rogan, Keselman, & Mendoza, 1979). Thus, prudent researchers need to be careful about mechanical use of this test. It makes sense to check the underlying normality of the distributions of data (and the transformed variables tracking the differences between each pair of repeated measures) before relying on the Mauchly test.

If sphericity is not an assumption the researcher is prepared to make, a novice researcher probably should suspend analyses until help is found to determine what elements are troubling the data set. For an experienced researcher, it may make sense to use a statistical adjustment. A value known as epsilon (ε) is identified (Greenhouse & Geisser, 1959)[7] under these circumstances. Epsilon may range from 0 to 1. If sphericity is perfectly met, $\varepsilon = 1.0$. If sphericity is not met, then epsilon will be lower than 1.0 (the worst case would be a number equal to $\frac{1}{k-1}$, where k is the number of treatment conditions or repeated measures). To adjust the analysis of variance, the Greenhouse-Geisser estimator multiplies the numerator and denominator degrees of freedom by ε. Though this approach tends to keep true Type I error close to the announced alpha risk (Collier et al., 1967; Stoloff, 1967), when the value of ε is greater than .7, an alternative called the Huynh-Feldt estimator is recommended (Huynh, 1978; Huynh & Feldt, 1976).[8] Otherwise, the Greenhouse-Geisser approach will tend to

---

[7] $\hat{\varepsilon}$ is computed as

$$\hat{\varepsilon} = \frac{k^2\,(\bar{s}_{ii} - \bar{s})^2}{(k-1)\left(\sum\sum s_{ijs}^2 - 2k\sum_i \bar{s}_i^2 + k^2\bar{s}^2\right)},$$

where

$\bar{s}$ is the mean of all the entries in the covariance matrix **S**,

$\bar{s}_{ii}$ is the mean of entries on the main diagonal of **S**,

$\bar{s}_i$ is the mean of all entries in row i of **S**, and

$\bar{s}_{ij}$ is the ijth entry of **S**.

[8] When the value of ε is greater than .7, Huynh and Feldt (1976) recommend that the computation of epsilon be adjusted as follows:

$$\bar{\varepsilon} = \frac{n(i-1)\hat{\varepsilon} - 2}{(i-1)[(n-1) - i - 1)\hat{\varepsilon}]}.$$

produce unacceptably conservative tests of statistical significance. For this reason, the method sometimes is known as the "Greenhouse-Geisser conservative *F* test" or sometimes just "the conservative *F* test." Unfortunately, using the Huynh-Feldt estimate of epsilon often may overestimate ε and, hence, will tend to increase the chances of committing Type I error. Hence, if sphericity is a problem, researchers are advised to look at *both* the Greenhouse-Geisser and the Huynh-Feldt formulations and to assume that the true *F* ratio lies somewhere between the two values. Kirk (1982, p. 261) recommends a three-step approach:

1. Check to see if the *F* statistic would be significant if sphericity were assumed (if not, stop the analysis, because there is little reason to believe that a statistically significant difference is present).

2. Use the Greenhouse-Geisser conservative *F* test (if the test is significant, claim a difference and stop).

3. Use the Huynh-Feldt method to see if the observed test statistic exceeds the critical value (if so, a difference is claimed as statistically significant).

Table 9.4 shows an example of such an analysis.

Table 9.4 Repeated Measures Example

A researcher was interested in learning if students changed their attitudes toward a topic that simply was mentioned but not explained in another speech. The attitudes about the "Simpson Environmental Protection Act" were taken for six people at 2, 4, and 6 weeks following their hearing a speech that mentioned the nonexistent act.

Subject	*Time 1*	*Time 2*	*Time 3*	*Mean*
1	4	5	8	5.67
2	1	2	5	2.67
3	2	3	5	3.33
4	5	4	8	5.67
5	1	3	6	3.33
6	5	7	10	7.33
Mean	3	4	7	Grand mean = 4.67

Computation:

$$\text{MS}_{\text{Time(a.k.a. within subjects)}}(\text{columns}) = n * s^2_{\overline{X}_k} = 6 * 4.33 = 26,$$

where n is the number of events measured at each time

$$\text{MS}_{(\text{Subjects})}(\text{rows}) = k * s^2_{\overline{X}_{j.}} = 3 * 3.33 = 10,$$

where k is the number of repeated measures (times each event is measured)

$$\text{Interaction SS}: \sum\left(\overline{X}_{jk} - \overline{X}_{j.} - \overline{X}_{.k} + \overline{\overline{X}}\right)^2$$
$$= [(4 - 5.67 - 3 + 4.67)^2 + \ldots (10 - 7.33 - 7 + 4.67)^2] = 4$$

Degrees of freedom for the time (a.k.a. within subjects) (columns) effect are the number of levels minus one.

Degrees of freedom for the subject (rows) effect are the number of levels minus one.

Degrees of freedom for the interaction effect are degrees of freedom for each of the main effects in the interaction multiplied by each other.

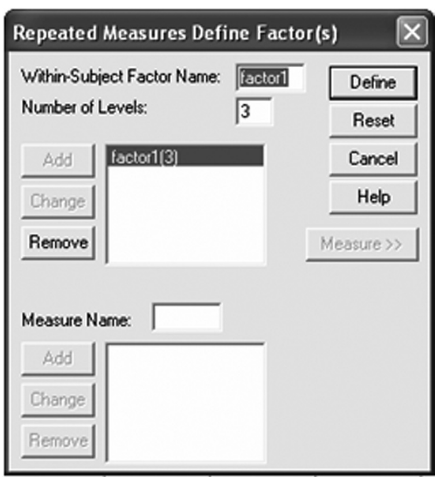
Because there is only one event in each "cell," within-groups variation is impossible to compute, but as with other mixed-effects designs, researchers may use the interaction mean square to get the job done. The following ANOVA table shows the results of the repeated measures analysis. As can be seen, the results show that the within-subjects "column" produced a statistically significant effect. Thus, the treatments evinced different effects on the dependent variable. Of course, because the subjects were different, it is not surprising that they showed some differences from each other as well, even though it is a matter of secondary interest here.

Source	*Sums of Squares*	*d.f.*	*Mean Square*	*F*
Columns (time)	52	2	26	65
Rows (subjects)	50	5	10	25
Interaction error	4	10	0.4	

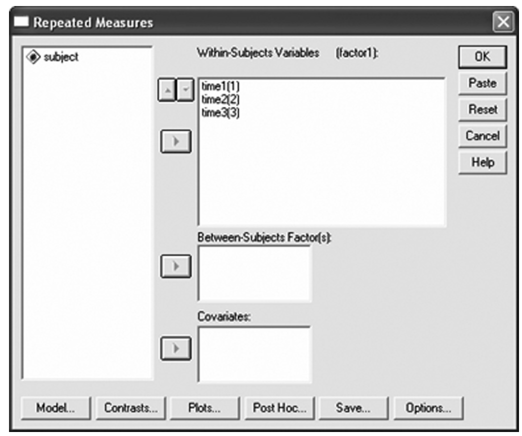
These results show a repeated measures analysis assuming sphericity and with no adjustments made. Given the complexity involved in completing Mauchly's *W* by hand, the application of this test to these data is covered in this chapter's section on using SPSS. For these data, the sphericity assumption was tenable and the Mauchly's test was not statistically significant.

Using SPSS for Repeated Measures ANOVA

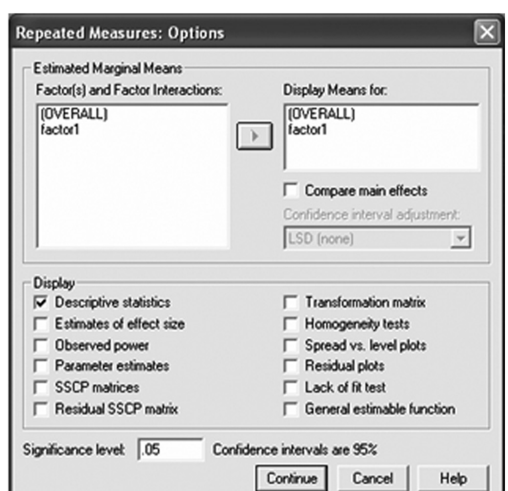
In the repeated measures design, the repeated measure sequence on which participants differ from one occasion to another is called the "within-subject factor." To run a repeated measures ANOVA, the researcher selects *General Linear Model* from the *Analyze* menu. Then, the researcher selects *Repeated Measures.* . . . In the *Repeated Measures Define Factor(s)* dialog box, the researcher assigns a name to the "Within-Subject Factor" on which the repeated measures were taken. As a default, the name "factor1" is used as a starting point. The researcher specifies the "Number of Levels," or times that the repeated measure was taken. The researcher clicks on the *Add* to include this repeated measure variable in analyses.



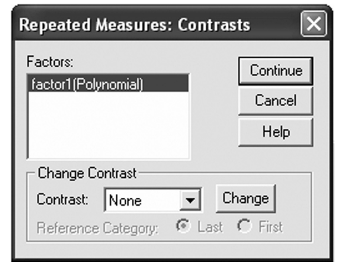
Then, the researcher clicks on the *Define* button to list the variables that have been used to identify each of the measurement occasions. For instance, in this case, the researcher measured the dependent variable three times for each person. The first time the variable was identified as "time1," the second time was called "time2," and the third time was "time3." In the *Repeated Measures* dialog box on the top of the right column, the researcher highlights the three variable "times" and then transfers them to the field marked "Within-Subject Variables (factor1):". Because the term "Subject" is simply a variable identifying the code number for the participants, it is not treated as another fixed variable.



Clicking the *Options. . .* button offers analysis tools of interest. The mean effects to be reported are highlighted and transferred to the field "Display Means for:", and the box to display descriptive statistics is checked.



After clicking the *Continue* button, the researcher clicks the *Contrasts* button. In the dialog box, "factor1" may be selected for a contrast. Because there are three levels, a linear and a quadratic effect could be examined. Afterward, the researcher would click *Continue* and then *OK*.



The first set of results shows the means, standard deviations, and sample size for each dependent variable scored at each time.

Descriptive Statistics

	Mean	Std. Deviation	N
TIME1	3.00	1.90	6
TIME2	4.00	1.79	6
TIME3	7.00	2.00	6

This output also includes the results of multivariate tests of significance. In this case, all tests have the same *F* ratios. As will be seen below, these *F* ratios are identical to the *F* ratio found for the test of the linear effect of the within-subjects effect. The multivariate option is used when there are additional main effects to test and/or when the assumption of sphericity is not tenable.

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
FACTOR1	Pillai's Trace	.981	102.000[a]	2.000	4.000	.000
	Wilks' Lambda	.019	102.000[a]	2.000	4.000	.000
	Hotelling's Trace	51.000	102.000[a]	2.000	4.000	.000
	Roy's Largest Root	51.000	102.000[a]	2.000	4.000	.000

a. Exact statistic
b. Design: Intercept
Within Subjects Design: FACTOR1

As can be seen here, the Mauchly's *W* was not statistically significant. Thus, the sphericity assumption was tenable, and the univariate repeated measures approach could be retained.

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
FACTOR1	.667	1.622	2	.444	.750	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subject Effect table.

b. Design: Intercept
Within Subjects Design: FACTOR1

The major test of the repeated measures ANOVA is reported on page 243. Because the sphericity assumption was reasonable, the standard ANOVA was employed and no adjusted tests were required. The "sphericity assumed" test showed a statistically significant *F* ratio. To identify the location of the differences, the researcher called for a trend analysis for the means, as reported on the "Tests of Within-Subject Contrasts" found on page 244.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FACTOR1	Sphericity Assumed	52.000	2	26.000	65.000	.000
	Greenhouse-Geisser	52.000	1.500	34.667	65.000	.000
	Huynh-Feldt	52.000	2.000	26.000	65.000	.000
	Lower-bound	52.000	1.000	52.000	65.000	.000
Error(FACTOR1)	Sphericity Assumed	4.000	10	.400		
	Greenhouse-Geisser	4.000	7.500	.533		
	Huynh-Feldt	4.000	10.000	.400		
	Lower-bound	4.000	5.000	.800		

Tests of Within-Subjects Contrasts

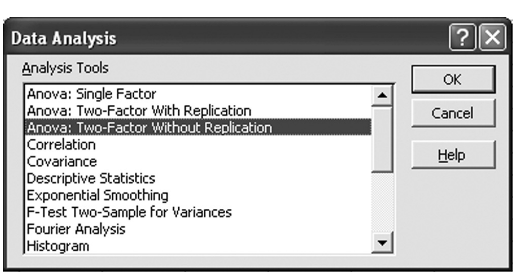
Measure: MEASURE_1

Source	FACTOR1	Type III Sum of Squares	df	Mean Square	F	Sig.
FACTOR1	Linear	48.000	1	48.000	120.000	.000
	Quadratic	4.000	1	4.000	10.000	.025
Error (FACTOR1)	Linear	2.000	5	.400		
	Quadratic	2.000	5	.400		

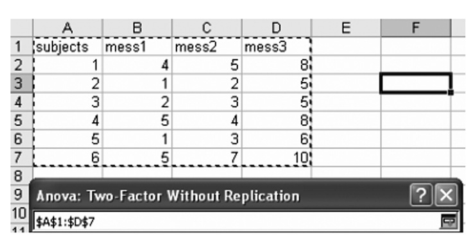
As shown here, a linear and a nonlinear (quadratic) effect were found. Thus, the presence of statistically significant effects when using higher-order polynomials would reveal that the best fit to the means is a curvilinear one. This effect is revealed by examining the means, which showed that the third condition was substantially above (higher than) the remaining means.

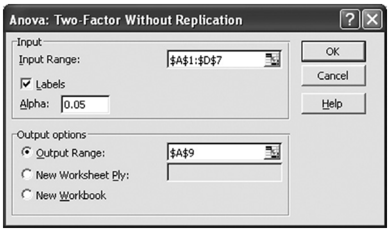
Using Excel for Repeated Measures ANOVA

To complete a repeated measures design, the researcher selects *Data Analysis* from the *Tools* menu and then chooses "Anova: Two-Factor Without Replication." After making this selection, the researcher clicks the *OK* button.



The data on the spreadsheet must be arranged in a particular way. The occasions for measuring the dependent variable are placed in columns, and the participants from whom repeated measures are taken are arranged as separate rows. The researcher selects the data by highlighting the cells in which the data are located.





Following the selection, the researcher clicks [button] to return to the dialog box. In the dialog box, the researcher must specify a desired alpha risk for use in significance testing (.05 in this case). In addition, the researcher specifies a location for the output (starting at cell A9 in this case) and then clicks *OK*.

The output appears as shown below. The analysis assumes sphericity but does not test for it. In addition, the alternative significance testing tools are not included. Overall, however, the major results are that a statistically significant difference was found from the repeated measures. Because this matter is the chief independent variable, it is the critical element for the researcher's hypotheses.

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
1	3	17	5.666667	4.33333
2	3	8	2.666667	4.33333
3	3	10	3.333333	2.33333
4	3	17	5.666667	4.33333
5	3	10	3.333333	6.33333
6	3	22	7.333333	6.33333
mess1	6	18	3	3.6
mess2	6	24	4	3.2
mess3	6	42	7	4

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	50	5	10	25	2.37771E-05	3.325837
Columns	52	2	26	65	1.85934E-06	4.102816
Error	4	10	0.4			
Total	106	17				