14 One-Way Analysis of Variance

For the model in Chapter 12 the dependent variable is caused by only the residual variable; that is,

$$\text{dependent variable} = \text{constant} + \text{residual variable}$$

In Chapter 13 added to the model is an independent variable, as follows:

$$\text{dependent variable} = \text{constant} + \text{effect of the independent variable} + \text{residual variable}$$

In Chapter 13 the independent variable is limited to a dichotomy, so only two groups can be compared. In this chapter the independent variable remains a nominal variable, but it may have more than two levels.[1]

The statistical technique used to analyze the model in which nominal independent variables affect a dependent variable measured at the interval level is called *analysis of variance*, commonly referred to as *ANOVA*. If there is a single nominal independent variable, the technique is called *one-way analysis of variance*. Analysis of variance is the most commonly used data analysis technique in psychology and is also commonly used in education, biology, and engineering. It represents not only a statistical test but also a way of thinking about research. In fact, one social psychologist, Harold Kelley, has suggested that persons in everyday life use something like analysis of variance to understand social reality.

The term "analysis of variance" is potentially confusing because the analysis of *variance* tests hypotheses about *means*. It must be remembered that an analysis of variability provides information about the means.

[1]The independent variable in ANOVA can be at the ordinal or interval level of measurement. The key requirement is that the variable have discrete groups. Because a nominal variable is always discrete, it is convenient to say that the independent variable is nominal.

In Table 14.1 is a data set that will be used in this chapter. A psychologist is interested in the effects of various strategies to enhance memory. Subjects were asked to memorize a list of 15 words and were tested one week later. The psychologist seeks to compare four different types of instruction: imagery, story, person, and none. In the imagery condition, subjects were told to picture each word. In the story condition, they were told to make up a story using the words. In the person condition, they were asked to associate each word with a person that they know. The none condition was a control condition. The means of the four groups, each with a sample size of ten, are as follows:

Imagery	*Story*	*Person*	*None*
12.3	10.6	9.4	7.3

Can the differences between the means be explained by chance? By chance is meant that the four groups are random samples drawn from the same population. Even if the groups were drawn from the same population, the means would still differ because of sampling error. At issue is whether the differences between means can be explained by chance. For example, 12.3 words recalled is greater than 10.6, but this 1.7 difference might have happened by chance. Just how likely would a 1.7 difference arise by chance? This is the type of question that the analysis of variance answers.

The ANOVA Model

The independent variable, called a *factor* in ANOVA, is denoted as Instructions or more simply as factor I. Ordinarily the factor is given an appropriate descriptive name, such as Drug or Reinforcement Schedule. The

TABLE 14.1 **Number of Words Recalled out of a Maximum of 15 Under Four Conditions**

	Instruction Level (I)			
	Imagery	Story	Person	None
	12	10	12	6
	14	9	8	4
	15	10	7	12
	10	11	5	8
	12	10	11	9
	14	13	13	11
	15	10	12	4
	12	11	10	6
	10	13	7	7
	9	9	9	6
ΣX	123	106	94	73
$\bar{X}$	12.3	10.6	9.4	7.3

factor can be abbreviated using a single uppercase letter such as D for Drug or R for Reinforcement Schedule. Categories of the independent variable are referred to as levels or groups. For instance, the factor in Table 14.1 has four levels or groups.

The variance of the means can be computed. This variance is

$$\frac{12.3^2 + 10.6^2 + 9.4^2 + 7.3^2 - (12.3 + 10.6 + 9.4 + 7.3)^2/4}{3}$$

which equals 4.42. The variance of the means provides a quantitative index of how different the four means are. Recall that the variance of the means can be interpreted as one-half the average squared difference among all pairs of means.

The variance within each of the groups can also be computed. The respective variances for the four groups are 4.68, 2.04, 6.93, and 7.34. The average of these variances is 5.25. This is the average or pooled variance within groups and is designated by s_p^2 because it is analogous to the pooled variance in Chapter 13.

There are now two measures of variability: the variance of the means, which equals 4.42, and the variance pooled within groups, which equals 5.25. As discussed in Chapter 11, the variance of the mean is a function of sample size. The variance of the mean is σ^2/n, where σ^2 is the variance of the observations that are used to make up the mean and n is the sample size. Thus, as sample size increases, the means should be more tightly bunched. So because the variance of the means is an inverse function of sample size, the variance of the means should be corrected by multiplying by the sample size; that is, $ns_{\overline{X}}^2$, where $s_{\overline{X}}^2$ is the variance of the group means and n is the group size. This quantity provides a measure of the variability of group means corrected for sample size.

The complete model for one-way ANOVA is

$$\begin{matrix}\text{dependent} \\ \text{variable}\end{matrix} = \text{constant} + \begin{matrix}\text{effect of the} \\ \text{independent} \\ \text{variable}\end{matrix} + \begin{matrix}\text{residual} \\ \text{variable}\end{matrix}$$

In the restricted model the independent variable has no effect; thus

$$\begin{matrix}\text{dependent} \\ \text{variable}\end{matrix} = \text{constant} + \begin{matrix}\text{residual} \\ \text{variable}\end{matrix}$$

If the restricted model were true, then both $ns_{\overline{X}}^2$ and s_p^2 estimate the variance of the residual variable. Thus, if the independent variable has no effect, the value of $ns_{\overline{X}}^2$ and s_p^2 should be close together because they both estimate the variance of the residual variable. In the memory example, $ns_{\overline{X}}^2$ equals 44.2, which is substantially larger than s_p^2, which equals 5.25. The two statistics do not appear to be estimating the same variance.

If the restricted model, which has no effect of the independent variable, is false, then only s_p^2 estimates the variance of the residual variable. The variance of sample means adjusted for sample size estimates the variance of the residual variable plus the variance of the population means times the sample size. So, if the independent variable has an effect, the variance of the means tends to be greater than the pooled variance, and this is the case in the example. So, by computing two variances—one the variance within groups and the other the variance of group means—hypotheses about means can be tested. This is the fundamental logic of analysis of variance.

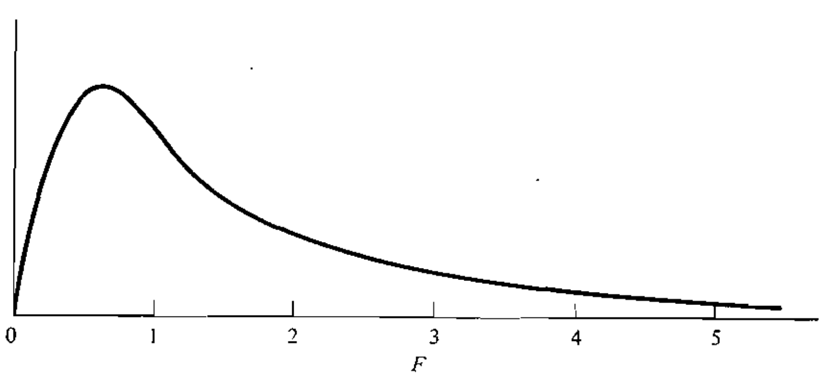
If an independent variable called factor A affects the dependent variable, then the means for the various levels of factor A should differ from the grand mean in the population. The null hypothesis of one-way analysis of variance is that the population means of the k groups are all equal to each other.

The restricted model in one-way analysis of variance is that the independent variable has no effect on the dependent variable. Under the restricted model, the variance of the group means times the number of persons in each group and the pooled variance within groups both estimate the residual variance. Given assumptions that underlie the restricted model (to be discussed later in this chapter), the ratio of these two estimates is distributed as F. If the restricted model is false, the value of F should be large.

The F distribution has a lower bound of zero, an upper bound of positive infinity, and a mode near one. An example of F is presented in Figure 14.1. There is not one F distribution but a family of distributions. The F distribution has two parameters. They are the degrees of freedom on the numerator and degrees of freedom on the denominator. They are symbolized by df_n and df_d, respectively. An F test statistic is written as $F(df_n, df_d)$. How to calculate the degrees of freedom for one-way analysis of variance will be explained later.

To find the p value of a test statistic distributed as F, the appropriate degrees of freedom for the numerator are located in the top row of Appendix E of the appropriate page, and the degrees of freedom in the denominator are

FIGURE 14.1 **Example of the *F* distribution.**



located in the first column. For a value to be significant, it must equal or exceed the tabled value, which is always some value greater than one. The significance level is obtained by choosing the largest value equaled or exceeded. In Appendix E are four significance levels: .10, .05, .01, and .001. (The F distribution was also presented in Chapter 11.)

Estimation and Testing

The term X_{ij} symbolizes the score for person i in group j. The first subscript refers to person and the second to group or level of the independent variable. There are n_j persons in group j, and there is a total of k groups. If for each group the group sizes are the same, the group size is denoted as n. The total number of scores is symbolized by N. So for the example in Table 14.1, n is 10, k is 4, and N is 40.

The sum of all the scores is symbolized by $\Sigma\Sigma X_{ij}$. The double summation signs indicate that scores are first summed in a group and then these group sums are added together. The sum of all the scores in a given group, say j, is symbolized by ΣX_{ij}. So the double summation indicates the sum of all the scores and the single sum indicates the sum of scores in a group.

The mean for group j, symbolized by $\bar{X}_{.j}$, is

$$\bar{X}_{.j} = \frac{\sum X_{ij}}{n_j}$$

The sum of the groups sizes or Σn_j is denoted as N. If group sizes are equal, N equals nk or group size times the number of groups. The mean of the means or the grand mean $\bar{X}_{..}$ equals

$$\bar{X}_{..} = \frac{\sum\sum X_{ij}}{N}$$

The dot notation indicates that the mean is computed across that subscript.

Alternatively, the grand mean can be computed as a weighted mean of the group means. The grand mean is computed by

$$\bar{X}_{..} = \frac{\sum n_j \bar{X}_{.j}}{N}$$

If the sample sizes are equal, the set of k means can be averaged to compute the grand mean, as follows:

$$\bar{X}_{..} = \frac{\sum \bar{X}_{.j}}{k}$$

It is useful to compute the grand mean both ways as a check on the computations.

To measure the constant in the complete model, the grand mean is used. To measure the effect of the independent variable, use the following:

$$\bar{X}_j - \bar{X}_{..}$$

that is, the mean for level j minus the grand mean. To compute the residual score for person i in group j, use:

$$X_{ij} - \bar{X}_j$$

that is, the score for that person minus the group mean. The estimated residual score is used to test assumptions of ANOVA that are presented later in this chapter.

The Analysis of Variance Table

Analysis of variance has a whole set of special terms, which are summarized in a table, such as Table 14.2. There are three rows in a one-way ANOVA table. The label for the top row is groups or in this case factor A. This line of the analysis of variance table represents variance attributable to groups. The term is sometimes referred to as the between-groups term. The second line represents the variation of subjects within groups. It is usually abbreviated as S/A, where the slash indicates within. This term is sometimes referred to as the *within-groups term* or *error term*. Subjects are said to be nested within levels of factor A. That is, each person is a member of one and only one group. The last line represents the total variation. Total is commonly abbreviated as TOT.

There are four columns to an analysis of variance table. The first column is for sum of squares. It can be abbreviated as SS. The sum of squares for groups or SS_A equals

TABLE 14.2 **Analysis of Variance Table**

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Factor A	SS_A	$k - 1$	$\frac{SS_A}{df_A}$	$\frac{MS_A}{MS_{S/A}}$
Subjects within groups (S/A)	$SS_{S/A}$	$N - k$	$\frac{SS_{S/A}}{df_{S/A}}$	
Total (TOT)	SS_{TOT}	$N - 1$		

$$SS_A = \sum n_j(\bar{X}_{.j} - \bar{X}_{..})^2$$

For this sum of squares, the term that is squared is the estimate of the treatment effect.

The sum of squares for subjects within levels of factor A, or $SS_{S/A}$, is as follows:

$$SS_{S/A} = \sum\sum(X_{ij} - \bar{X}_{.j})^2$$

Each score has its group mean subtracted from it, the difference is squared, and then all the squared differences are summed across the entire set of scores. As stated earlier, the quantity $X_{ij} - \bar{X}_{.j}$ is the estimate of the residual score for person i in group j.

The sum of squares total or SS_{TOT} is as follows,

$$SS_{TOT} = \sum\sum(X_{ij} - \bar{X}_{..})^2$$

again where $\bar{X}_{..} = \Sigma\Sigma X_{ij}/N$. It can be shown that

$$SS_{TOT} = SS_A + SS_{S/A}$$

and so the $SS_{S/A}$ can be obtained by subtraction as follows:

$$SS_{S/A} = SS_{TOT} - SS_A$$

Ordinarily the $SS_{S/A}$ is computed indirectly by subtracting the SS_A from the SS_{TOT}.

The second column contains the degrees of freedom or *df*. The total degrees of freedom are

$$df_{TOT} = N - 1$$

or the total sample size less one. The degrees of freedom for factor A are:

$$df_A = k - 1$$

or the number of groups less one. The degrees of freedom for *S/A* are

$$df_{S/A} = N - k$$

or the total sample size minus the number of groups.

It is helpful to remember that degrees of freedom partition in the same way as sums of squares. So just as $SS_{S/A} = SS_{TOT} - SS_A$, it is also true that

$$df_{S/A} = df_{TOT} - df_A$$

Any partitioning of the sums of squares can also be done to the degrees of freedom.

The *mean square* equals the sum of squares for the line divided by the degrees of freedom for the line. So to compute MS_A, the SS_A is divided by

df_A; and to compute $MS_{S/A}$, the $SS_{S/A}$ is divided by $df_{S/A}$. Usually the MS_{TOT} is not computed, but it would equal the variability of the observations ignoring the independent variable.

The MS_A is the variability of the groups means corrected for sample size. The $MS_{S/A}$ is the pooled variance within levels of factor A. Both mean squares estimate the residual variance given that the restricted model is true. The final column in the analysis of variance table is the F test statistic, which equals

$$F(k\text{–}1,\ N\text{–}k) = \frac{MS_A}{MS_{S/A}}$$

The degrees of freedom for this F test on the numerator or df_n are $k - 1$ and the degrees of freedom on the denominator df_d are $N - k$. The F test evaluates the restricted model. It essentially compares the variability in the group means to the variability of scores within groups. If the restricted model is false, then F tends to be large. If the F is statistically significant, the restricted model is rejected, and the complete model in which the independent variable, factor A, affects the dependent variable is preferred.

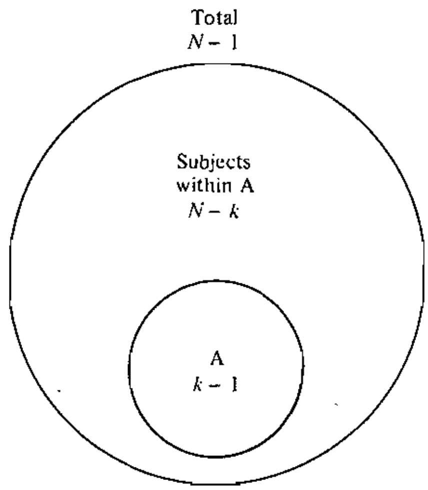
In Table 14.3 is the analysis of variance table for the data in Table 14.1. The F ratio of 8.42 is statistically significant at the .001 level. It indicates that "Instruction" affects the dependent variable. Thus, subjects recalled different amounts of material in the four conditions.

Most of the effort in analysis of variance involves the computation of sums of squares. For every analysis at least one sum of squares is computed from others. For instance, in one-way ANOVA the sum of squares within groups or $SS_{S/A}$ equals the sum of squares total, SS_{TOT}, minus the sum of squares for the independent variable, SS_A. It is helpful to draw a circle diagram that illustrates the partitioning of the sum of squares, as in Figure 14.2. The complete circle represents the sum of squares total. The circle inside the circle represents the sum of squares for factor A, and the remainder represents the sum of squares for persons within levels of factor A. For more complex designs, these diagrams can be especially helpful.

TABLE 14.3 **Analysis of Variance Table for the Data in Table 14.1**

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Instruction (I)	132.6	3	44.20	8.42
Subjects within I (S/I)	189.0	36	5.25	
Total	321.6	39		

FIGURE 14.2 **Circle diagram for one-way analysis of variance.**



Computational Formulas for the Sum of Squares

The computation of various ANOVA terms is discussed, first for the case when groups are equal in size and second for the case when group sizes are unequal.

For the computation of the sum of squares, it is necessary to define various totals. They are the grand total

$$T_{..} = \sum\sum X_{ij}$$

and the group total for group j

$$T_{.j} = \sum X_{ij}$$

For the example in Table 14.1, the group totals are 123, 106, 94, and 73. The grand total, which is the sum of the group totals, equals 396.

To compute the F test to evaluate the relative plausibility of the complete and the restricted models, both SS_A and $SS_{S/A}$ must be computed. First the *correction term for the mean, C,* is computed as follows:

$$C = \frac{(\sum\sum X_{ij})^2}{N}$$

or using totals, $C = T_{..}^2/N$. In words, the correction term for the mean is the square of the sum of all the observations, divided by the total number of observations in the study. It is called a correction term for the mean because it tends to be large if the mean is large. The sum of all the observations for the memory example is 396, and so the correction term for the mean is $396^2/40 =$ 3920.4.

The sum of squares for factor A, or SS_A, is

$$SS_A = \frac{\sum T_{.j}^2}{n} - C$$

where again C equals $T_{..}^2/N$.

The $SS_{S/A}$ is computed indirectly, as follows:

$$SS_{S/A} = SS_{TOT} - SS_A$$

The sum of squares total equals

$$SS_{TOT} = \sum\sum X_{ij}^2 - C$$

That is, it equals the sum of the square of every observation, minus C.

If group sizes are unequal, the SS_A equals

$$SS_A = \sum \frac{T_{.j}^2}{n_j} - C$$

So the group total is squared and divided by the group's n. These are then summed across the k groups, and the mean's correction term C is subtracted.

Test of the Constant

As shown in Chapter 12, a researcher may have an a priori hypothesis concerning the constant in the model. For instance, if a recognition memory test is used, the constant should equal some a priori value if subjects were guessing. It is possible to test a restricted model that the mean equals some value, within a one-way ANOVA.

The a priori constant is denoted as M. The sums of squares for the constant are $N(\bar{X}_{..} - M)^2$. (Note that if M is zero, this quantity equals C, the correction term for the mean.) The sum of squares for the constant divided by the $MS_{S/A}$ is an F test with 1 and $N - k$ degrees of freedom. If the F is statistically significant, the null hypothesis that the population mean equals M is rejected.

As an example, suppose it is desired to test the null hypothesis that the constant in Table 14.1 is 10.0. The grand mean equals 9.9. The sum of squares for the constant is $40(10.0 - 9.9)^2$, which equals .40. From Table 14.3 the $MS_{S/A}$ is 5.25 and so the F is .40/5.25 or .08. Because the F is not significant, the constant is not significantly different from 10.0.

Assumptions and Interpretation

The assumptions of analysis of variance are basically the same as those of the t test discussed in the previous chapter. These assumptions all refer to the residual variable and they are (a) normal distribution, (b) equal variability, and (c) independence of observations.

Because these assumptions are extensively discussed in the previous chapter, they will not be reviewed here except to repeat that the independence assumption is the one assumption that must be carefully scrutinized. The reader should consult Chapter 13 for a discussion of these assumptions.

If the group means differ significantly, the meaning of those differences depends on design considerations. If persons are randomly assigned to groups, then the significant effect can be attributed to the independent variable. Without random assignment it is not clear what causes what. For instance, consider a study on the effects of jogging. Three groups are formed: nonjoggers, joggers, and dedicated joggers. The dependent variable is weight. Suppose it is found that persons who jog more often weigh less. We cannot conclude that jogging causes a loss in weight; the truth might be that overweight persons simply do not wish to jog. That is, it is the dependent variable that causes the independent variable and not vice versa. Without random assignment of persons to groups, it cannot be unequivocally concluded that the independent variable causes the dependent variable.

Power and Measure of Effect Size

Recall that the power of a test is the probability of rejecting the restricted model given that the restricted model is false. To evaluate the power the following factors must be considered: First, how different are the population means from one another? The more different they are, the greater the power. So the more truly different the groups are, the more likely the restricted model will be rejected. Second, how many persons are in each group? The more persons in each group, the greater the power. Third, how great is the variability within treatment groups? The more similar persons within the same group are, the greater the power. All three of these factors were considered in more detail in the previous chapter.

The most common measure of effect size for one-way analysis of variance is called *omega squared* or, as it is commonly symbolized, ω^2. Its computational formula is

$$\omega^2 = \frac{\text{SS}_A - (k - 1)\text{MS}_{S/A}}{\text{SS}_{\text{TOT}} + \text{MS}_{S/A}}$$

Omega squared can be interpreted as the proportion of variance in the dependent variable that is explained by the independent variable. So, like a correlation coefficient, the upper limit of omega squared is 1.00. If the estimated value of omega squared is less than zero, then omega squared is set to zero. From the values in Table 14.3,

$$\omega^2 = \frac{132.6 - 3(5.25)}{321.6 + 5.25} = .36$$

It should be noted that omega squared is in squared units. If the standards for large, medium, and small that were set for the d in the previous chapter are used for ω^2, a large ω^2 is .40, a medium value is .20, and a small value is .04.

Contrasts

The complete model postulates some effect of the independent variable on the dependent variable. It does not, however, explicitly predict exactly how the means differ, but only that at least two groups have different population means. The researcher may have a clear idea about exactly how the means differ. Consider some specific examples.

1. A researcher is interested in examining the effect of the day of the week on absenteeism. She believes that absenteeism is higher on Monday than any other day of the week. She conducts a one-way analysis of variance using day of the week as the independent variable.
2. A researcher believes that more study time will improve performance on an examination. She creates four groups who study zero, one, two, and three hours. She then measures performance on an examination. She analyzes her data by a one-way analysis of variance.
3. A researcher creates four groups to investigate the effects of smoking and stress on blood pressure. The four groups are

    I: No smoking, no stress
    II: Smoking, no stress
    III: No smoking, stress
    IV: Smoking, stress

    He analyzes his data by a one-way analysis of variance, although he wished to compare those who smoked with those who did not.

In all of these examples the researchers had a specific hypothesis about the patterning of the means. However, they did not explicitly test for this pattern. This failure to make such a test can lead to mistaken conclusions. If the F is significant, it does not imply that the researcher's hypothesis is true. Conversely, if the F is not significant, it does not imply that the researcher's hypothesis is false. The overall F test does not directly test the researcher's specific hypothesis. It only evaluates the very general null hypothesis that all the population means are equal to each other.

When there are three or more groups and an explicit hypothesis about how the means differ, a contrast can be used to test that hypothesis. A *contrast* is a set of weights assigned to each level of the independent variable to evaluate an explicit hypothesis. These weights that tap the hypothesis of interest must sum to zero.

A *contrast* is a set of numbers, each of which is paired with one level of the independent variable. To determine the particular contrast weights, the null hypothesis must be explicitly stated. For instance, for the researcher who is interested in Monday absenteeism, the null hypothesis is that Monday does not differ from the other four days of the week. Algebraically, this null hypothesis can be expressed as

$$\mu_{MO} = \frac{\mu_{TU} + \mu_{WE} + \mu_{TH} + \mu_{FR}}{4.0}$$

(The μ terms are the population means.) The terms can be rearranged and all put on the right-hand side:

$$0 = \mu_{MO} - .25\mu_{TU} - .25\mu_{WE} - .25\mu_{TH} - .25\mu_{FR}$$

The contrast weights are the numbers that multiply the means; that is,

Monday	*Tuesday*	*Wednesday*	*Thursday*	*Friday*
1.0	–.25	–.25	–.25	–.25

Note that the five contrast weights sum to zero, as they must.

Sometimes the researcher has quantitative values attached to the levels of the independent variable. To derive the contrast weights, the quantitative values are averaged. The contrast weight for a given level equals the quantitative value minus this average. So, for the second example, persons study zero, one, two, and three hours. The average of the four numbers is 1.5. The contrast weights are

Number of Hours Studied			
0	1	2	3
–1.5	–.5	.5	1.5

Again, the sum of the contrast weights is zero.

For the final example, the researcher seeks to compare smokers (groups II and IV) with nonsmokers (groups I and III). The null hypothesis is $\mu_{I} + \mu_{III} = \mu_{II} + \mu_{IV}$. Putting all the terms on the right-hand side yields

$$0 = \mu_{II} + \mu_{IV} - \mu_{I} - \mu_{III}$$

The contrast weights are then

Group			
I	II	III	IV
–1	+1	–1	+1

To test a contrast, its sum of squares must be computed. The contrast value for the jth mean is denoted as p_j. Given an equal number of observations in each group (that is, n_j is constant), the sum of squares for a contrast equals

$$\frac{(\sum_j p_j T_{.j})^2}{n \sum p_j^2}$$

The degrees of freedom for a contrast are equal to one. So, the sum of squares and the mean square are the same. To test the contrast, the mean square for the contrast is divided by the $MS_{S/A}$. Given the null hypothesis, the ratio is distributed as F with 1 and $N - k$ degrees of freedom. Formulas for the sum of squares for a contrast with unequal group sizes are given in more advanced texts (Myers, 1979; Winer, 1971).

A contrast is a dummy variable. To each level of a nominal variable (the independent variable) numerical values (contrast weights) are attached. This dummy variable can be viewed as the predictor variable in a regression equation in which the dependent variable serves as the criterion. If such a dummy variable is created and the regression coefficient is computed and tested using the method described in Chapter 16, the p value is the same as the one-way ANOVA test of the contrast.

Post Hoc Tests

The use of contrasts to test an explicit hypothesis is called an *a priori test*. If the researcher has no explicit hypothesis about the patterning of the means, he or she can perform what is called a *post hoc test* to determine how it is that the means differ. There are many, many different ways of performing post hoc tests and there is no clear consensus about which is the best technique. The Tukey least significant difference (lsd) test is presented because it is relatively easy to compute. The reader is referred to more advanced texts for descriptions of other post hoc test procedures (Myers, 1979; Winer, 1971).

The Tukey lsd test is called a *protected* test. It can only be done if the F test is significant. Assuming that it is, all possible pairs of means are compared. Each pair of means is tested using the formula

$$\text{Tukey lsd} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{MS_{S/A}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

where $\bar{X}_1$ and $\bar{X}_2$ are the sample means and n_1 and n_2 are the respective sample sizes. The formula for the Tukey lsd test is a t test of the difference between means (see the previous chapter) with $MS_{S/A}$ substituted for s_p^2. To determine whether the difference is statistically significant, the t distribution with $N - k$ degrees of freedom is used, where k is the number of groups.

As an example consider the data in Table 4.1. Because the F is significant, the Tukey lsd test can be used. Because there are four groups, there are six possible comparisons. The results are

Imagery with Story:	$t(36) = 1.659$
Imagery with Person:	$t(36) = 2.830$
Imagery with None:	$t(36) = 4.880$
Story with Person:	$t(36) = 1.711$
Story with None:	$t(36) = 3.220$
Person with None:	$t(36) = 2.049$

Using Appendix D, for a t with 36 degrees of freedom to be statistically significant at the .05 level of significance, it must be at least 2.030. So, four of the six comparisons are judged as significant.

When group sizes are equal, the smallest difference between means that would be significant—that is, the least (l) significant (s) difference (d)—equals

$$t(.05,\ df)\ \frac{\sqrt{2\text{MS}_{S/A}}}{\sqrt{n}}$$

where $t(.05,\ df)$ is the critical value for .05 significance with degrees of freedom of df. For the example, the lsd is

$$2.030\ \frac{\sqrt{2(5.25)}}{\sqrt{10}} = 2.080$$

Hence a difference between a pair of means must be at least 2.080 to be significantly different at the .05 level of significance.

Illustrations

Example 1

An experimenter seeks to compare the degree of comfort of two automobiles. Twenty different persons are asked to sit in one of the two automobiles. Comfort is measured on a ten-point scale, with higher numbers indicating greater comfort. The cars are designated as car 1 and car 2. The numbers are

Car 1: 8, 9, 7, 8, 7, 6, 9, 6, 7, 5

Car 2: 10, 9, 8, 10, 9, 8, 7, 9, 9, 8

The means for the two cars are 7.2 and 8.7, and the grand mean is 7.95. Car 2 is rated as more comfortable, but it must be determined whether the difference is statistically significant.

The grand total or $T_{..}$ equals 159, the group totals are 72 and 87, N is 20, n is 10, and k is 2. The correction term for the mean is $159^2/20$ or 1264.05. The sum of squares for cars is

$$\frac{72^2 + 87^2}{10} - 1264.05 = 11.25$$

The sum of each squared score is 1299. The sum of squares total is

$$1299 - 1264.05 = 34.95$$

By subtraction, the sum of squares for subjects within cars is

$$34.95 - 11.25 = 23.70$$

The analysis of variance table is presented in Table 14.4.

The F value of 8.54 is statistically significant at the .01 level. Car 2 is significantly preferred over car 1. The value of omega squared for this example is .27, which is between moderate and large.

If a t test had been done comparing the two groups, it would have been found that the t value is 2.92, which is the square root of 8.54. When there are only two groups in a one-way analysis of variance, the square root of F exactly equals the value that would be obtained in a two-group t test. So, an ANOVA with two levels is equivalent to the two-group t test described in the previous chapter. [In Chapter 11 it was pointed out that $t(df)^2 = F(1,df)$.]

Example 2

A sociologist studies the degree of satisfaction of workers. She wishes to compare the satisfaction with job of secretaries, janitors, and managers. She uses a 20-point scale to measure job satisfaction where higher numbers indicate greater satisfaction. Her results are

Secretaries: 12, 16, 15, 19, 14

Janitors: 17, 19, 14, 18

Managers: 20, 19, 18

This is an unequal n design because there are five secretaries, four janitors, and three managers. The grand mean is 16.75. The mean for secretaries is 15.2, for janitors 17.0, and for managers 19.0.

TABLE 14.4 **Analysis of Variance Table for Car Study**

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Car (C)	11.25	1	11.250	8.54
Subjects within C (S/C)	23.70	18	1.317	
Total	34.95	19		

The correction term for the mean is $201^2/12 = 3366.75$. The sum of squares for job is

$$\frac{76^2}{5} + \frac{68^2}{4} + \frac{57^2}{3} - 3366.75 = 27.45$$

The sum of all the squared scores is 3437 and so the sum of squares total is

$$3437 - 3366.75 = 70.25$$

The sum of squares for persons within jobs is $70.25 - 27.45 = 42.80$. The analysis of variance table is presented in Table 14.5. The F is not statistically significant at the .05 level even though omega squared has a moderately large value of .24. So there is no evidence that the workers differ in their satisfaction. The sample sizes are so small that the level of power is quite low even for a moderate effect size. A Tukey lsd post hoc test should not be done because the F is not significant.

Example 3

Three different types of psychotherapy are to be evaluated. A fourth group which is a control group is also set up. There are five persons in each group. The scores on an adjustment scale (the larger number, the greater the adjustment) are

Therapy Group I: 6, 7, 5, 7, 4

Therapy Group II: 4, 5, 6, 7, 4

Therapy Group III: 3, 5, 4, 3, 6

Control Group: 2, 4, 3, 2, 3,

The means for the three therapy groups are 5.8, 5.2, and 4.2, respectively. The mean for the control group is 2.8, and so the persons receiving psychotherapy are relatively more adjusted. The grand mean is 4.5.

TABLE 14.5 **Analysis of Variance Table for Job Satisfaction Study**

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Job (J)	27.45	2	13.725	2.89
Subjects within J (S/J)	42.80	9	4.756	
Total	70.25	11		

The correction term for the mean is $90^2/20 = 405$. The sum of squares for groups is

$$\frac{(29^2 + 26^2 + 21^2 + 14^2)}{5} - 405 = 25.8$$

The sum of squared observations is 454, and so the sum of squares total is

$$454 - 405 = 49$$

By subtraction (49.0 – 25.8) the sum of squares for subjects within groups is 23.2. The degrees of freedom for groups are $4 - 1 = 3$, and the degrees of freedom for subjects within groups are $20 - 4 = 16$. The analysis of variance table is presented in Table 14.6. The F for groups is statistically significant at the .01 level of significance, and omega squared is .43.

Although the F is highly significant, it is not known whether the difference is due to the therapy groups being higher than the control group. A contrast must be created to test this hypothesis. The contrast compares the average of the three therapy groups with that of the control group. This results in contrast weights of .33 for the three therapy groups and –1 for the control group. The sum of squares for the contrast is

$$\frac{[(.33)(29) + (.33)(26) + (.33)(21) + (-1.0)(14)]^2}{5(.33^2 + .33^2 + .33^2 + 1^2)} = 19.27$$

The F test of the contrast is

$$F(1,16) = \frac{19.27}{1.45} = 13.29$$

which is statistically significant at the .001 level.

Summary

One-way analysis of variance is a statistical procedure used to test the effect of a nominal variable on an interval variable. The independent variable is a

TABLE 14.6 **Analysis of Variance Table for the Therapy Study**

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Group (G)	25.8	3	8.60	5.93
Subjects within G (S/G)	23.2	16	1.45	
Total	49.0	19		

multilevel nominal variable. In the restricted model the independent variable has no effect on the dependent variable. Variation of scores is partitioned into two sources: between groups and within groups. For each source of variation its sum of squares and degrees of freedom are computed.

The degrees of freedom for the independent variable are $k - 1$, where k is the number of groups. The degrees of freedom for subjects within groups are $N - k$, where N is the total number of subjects in the study. The degrees of freedom for the total are $N - 1$. The sum of squares for subjects within groups equals the sum of squares total minus the sum of squares groups. A *mean square* equals a sum of squares divided by its degrees of freedom.

The sum of squares, degrees of freedom, and mean squares are summarized in an *analysis of variance table*. The fit of the restricted model is evaluated by an F test. The numerator of the test is the mean square for the independent variable and the denominator is the mean square subjects within levels of the independent variable.

The power of the F test in one-way ANOVA depends on the difference between the populations means, the group size, and the degree of similarity within groups. *Omega squared* is used to measure the proportion of variance in the dependent variable that is explained by the independent variable. *Contrasts* are used to test a priori hypotheses about the exact patterning of the means. The *Tukey lsd test* can be used to test differences between all possible pairs of means. This test is a *post hoc* test, which means that the researcher need not have any hypotheses about how the means differ.

In the next chapter the topic is two-way analysis of variance, involving two independent variables. Two-way ANOVA can be viewed as an extension of one-way ANOVA.

Problems

1. For the following significance levels and degrees of freedom determine the appropriate F value needed:

	Alpha	df_n	df_d
a.	.05	1	16
b.	.05	4	56
c.	.05	2	44
d.	.001	1	63
e.	.01	4	123
f.	.10	6	23
g.	.001	8	48
h.	.01	2	72

2. If there are five levels of factor A with sample sizes of 10, 9, 8, 6, and 10

in the five groups, complete the following analysis of variance table and give the significance level for F.

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F
Factor A	140.0			
Subjects within A (S/A)				
Total	520.0			

3. A psychologist is interested in the relationship between handedness and athletic ability. He measures the athletic ability of three groups of persons: left-handed, right-handed, and ambidextrous persons. His results are:

   Left-handed: 11, 13, 14, 13, 15

   Right-handed: 10, 8, 7, 10, 14

   Ambidextrous: 12, 8, 6, 11, 15

   Do a one-way analysis of variance to determine whether the groups differ significantly. Compute and interpret omega squared.

4. a. For the data in problem 3, do a Tukey lsd test of the difference between means.
   b. For the data in problem 3, test whether the constant is significantly different from 10.

5. A researcher seeks to measure the degree of allergic reaction to a drug. Is there a significant difference between the groups?

   Group I: 21, 19, 18, 13, 15, 20, 22, 25, 17, 17

   Group II: 12, 10, 20, 14, 18, 8, 12

   Use both analysis of variance and a t test to determine whether the groups significantly differ from one another.

6. Consider a one-way analysis of variance with five levels and twelve subjects in each level. Given that the means for the five groups are

   $$2.0, 3.2, 4.1, 5.2, 5.1$$

   create a contrast that compares the first two means with the second two. Compute the mean square for that contrast.

7. For the following analysis of variance table, compute and interpret the value of omega squared.

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F*
Factor T	55.33	3	18.44	7.41
Subjects within T (S/T)	139.44	56	2.49	
Total	194.77	59		

8. An investigator wishes to determine the effectiveness of three different treatments in relieving headache pain. The drugs to be studied are aspirin, acetaminophen, and a placebo. Ten different persons take one drug and rate their pain on a ten-point scale after three hours. The scores are as follows:

   Aspirin: 7, 6, 9, 5, 3, 5, 3, 2, 4, 2

   Acetaminophen: 5, 8, 6, 4, 7, 4, 6, 2, 3, 7

   Placebo: 9, 7, 8, 7, 5, 4, 6, 8, 3, 7

   Use analysis of variance to evaluate whether the groups differ. Compute and interpret omega squared.

9. For problem 8 create a contrast that compares the two medicated groups to the placebo group. Create a contrast that compares the aspirin group to the acetaminophen group. Test each contrast.

10. For the following ANOVA table compute and interpret omega squared.

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F*
Factor A	200	4	50	5
Subjects within A (S/A)	100	10	10	
Total	300	14		

11. For problem 10 answer the following questions.
    a. How many subjects are there in the study?
    b. How many levels of the independent variable are there?
    c. What are the degrees of freedom for the *F* test?
    d. May a Tukey lsd post hoc test be done on the means?

12. A researcher seeks to compare the marital satisfaction of women who have married for varying number of years. She finds the following (higher numbers, greater satisfaction).

One year: 56, 48, 57, 41

Two years: 63, 51, 65, 54

Ten years: 70, 61, 55, 58

a. Use one-way ANOVA to test the effect of length of marriage on satisfaction.
b. Test whether the average satisfaction score is significantly different from 55.

13. For problem 12 create a contrast to test the hypothesis that satisfaction increases (or decreases) for every year of marriage. Test the contrast by an F test.

14. A researcher seeks to determine if the maturity of a five-year-old's speech depends on the age of his or her partner. He pairs 15 individual five-year-old subjects with one of four types of partners: infant, five-year-old, twelve-year-old, or adult. The scores on a maturity scale are

    Infant: 3, 7, 5

    Five-year-old: 8, 11, 14

    Twelve-year-old: 11, 15, 18

    Adult: 14, 12, 17, 19, 15, 13

    Use one-way ANOVA to test whether the child adjusts his or her speech for different types of partners. Compute and interpret omega squared for the study.

15. A researcher seeks to determine the effect of a drug on the number of hours of sleep. Four different dosages are compared: none, 10 ml, 20 ml, and 30 ml. The results are

    None: 4, 6, 5, 8, 3, 2

    10 ml: 6, 8, 9, 6, 8, 4

    20 ml: 7, 9, 6, 5, 4

    30 ml: 9, 8, 7, 6

    Use one-way ANOVA to test the effect of the drug on the number of hours slept.

16. Conduct a Tukey lsd test for the data in problem 15.