Chapter twelve

Comparing Means Among More Than Two Samples: Analysis of Variance

D. Weisburd and C. Britt, *Statistics in Criminal Justice*, DOI 10.1007/978-1-4614-9170-5_12,

IN CHAPTER 11, we used the *t* distribution to test hypotheses about means from two independent samples. But what if we are interested in looking at more than two samples at a time? This is a common problem in criminology and criminal justice, where many important questions can be raised across a number of different samples. For example, race is a central concern in criminal justice and criminology, and often it does not make sense to restrict comparisons involving race to just two groups. Similarly, in many criminal justice studies, a number of interventions are compared simultaneously. In such studies, researchers want to compare not just two but three or more means in the context of one statistical test.

Analysis of variance (ANOVA) is a commonly used parametric test of statistical significance that allows the researcher to compare multiple groups on specific characteristics. ANOVA also provides an opportunity to introduce several important statistical concepts used in more complex types of analysis. In this chapter, we examine in this context the concepts of explained and unexplained variation and consider how they relate to the total variation found in a specific measure. This chapter also introduces a nonparametric test, the Kruskal-Wallis test, which can be used for comparisons across multiple groups when the assumptions underlying ANOVA are difficult to meet.

Analysis of Variance

Analysis of variance is based on a simple premise. As the differences between the means of samples become larger relative to the variability of scores within each sample, our confidence in making inferences grows. Why is this the case? Certainly it makes sense that the more the mean differs from one sample to another, the stronger the evidence supporting the position that there are differences between the population means. All else being equal, the larger the differences between the samples, the more confident we are likely to be in rejecting the position that the population means are equal.

But we are faced with a problem in making such an inference. How much confidence can we place in the observed means of our samples? As we have stated many times in this book, sample estimates vary from sample to sample. If the sample means are likely to vary considerably, then we want to be cautious in drawing strong conclusions from our study. If the sample means are not likely to vary greatly, we can have more confidence in conclusions drawn from them. Analysis of variance uses the variability within the observed samples to come to conclusions about this variability.

Suppose, for example, that you are examining two separate studies, each including three samples. In the first study, the scores are widely dispersed around the mean for each group. In contrast, in the second study, the scores are tightly clustered around the group means. If you take the variability you observe in these samples as an indication of the variability in the populations from which they were drawn, you are likely to have more confidence in estimates gained from the second study than from the first. Those estimates appear to be more stable, evidencing much less variability.

This is precisely the approach taken in analysis of variance. The variability *between* the groups studied is contrasted with the variability *within* these groups to produce a ratio:

$$\frac{\text{Variability between groups}}{\text{Variability within groups}}$$

The larger this ratio—the larger the differences between the groups relative to the variability within them—the more confidence we can have in a conclusion that the population means are not equal. When the ratio is smaller, meaning that the differences between the groups are small relative to the variability within them, we have less reason to conclude that differences exist in the populations to which we want to infer.

Developing Estimates of Variance Between and Within Groups

The first step in analysis of variance is to define the variability between and within the groups studied. To make this process more concrete, let's use a hypothetical study of depression among 12 prison inmates drawn from high-, moderate-, and low-security prisons (see Table 12.1).

Table 12.1 Depression Scores for 12 Prison Inmates Drawn from High-, Moderate-, and Low-Security Prisons

LOW SECURITY (GROUP 1)	MODERATE SECURITY (GROUP 2)	HIGH SECURITY (GROUP 3)
3	9	9
5	9	10
4	8	7
4	6	10
$\sum = 16$	$\sum = 32$	$\sum = 36$
$\bar{X} = 4$	$\bar{X} = 8$	$\bar{X} = 9$

To calculate the grand mean: $\bar{X}_g = \dfrac{\sum_{i=1}^{N} X_i}{N} = \dfrac{84}{12} = 7$

Between-group variability is measured by first subtracting the **grand mean,** or **overall mean,** of the three samples from the mean of each sample. This difference must then be adjusted to take into account the number of scores or observations in each sample. Equation 12.1 represents this process in mathematical language. The sample (or category) means are represented by $\bar{X}_c$, the overall mean (or grand mean) is represented by $\bar{X}_g$, N_c represents the number of scores or observations in the sample (or category), and $\sum_{c=1}^{k}$ tells us to sum the results from the first sample (or category) mean ($c = 1$) to the last sample (or category) mean ($c = k$).

$$\sum_{c=1}^{k} [N_c(\bar{X}_c - \bar{X}_g)]$$ **Equation 12.1**

As illustrated in Table 12.1, the overall, or grand, mean is found by adding up all the scores in the three samples and dividing by the total number of scores ($N = 12$). To calculate the amount of between-group variability for our example, we take just three quantities—the mean depression score of high-security inmates ($\bar{X} = 9$) minus the overall mean, the mean depression score of moderate-security inmates ($\bar{X} = 8$) minus the overall mean, and the mean depression score of low-security inmates ($\bar{X} = 4$) minus the overall mean—and multiply each by its sample size.

Within-group variability is identified by summing the difference between each subject's score and the mean for the sample in which the subject is found. In Equation 12.2, X_i represents a score from one of the three samples and, as before, $\overline{X}_c$ represents the mean for that sample. Here we sum from $i = 1$ to N, or from the first to the last score in the overall study.

$$\sum_{i=1}^{N} (X_i - \overline{X}_c)$$ **Equation 12.2**

For within-group variability, we have four calculations to carry out for each sample (group). For the first subject in the low-security prison sample, for example, we subtract from the subject's score of 3 the mean score of low-security inmates in the study (4). The same is done for each of the other three members of this sample. For the moderate-security sample, we repeat the process, starting with the first subject, with a score of 9, and using the mean of 8 for the group as a whole. The same is done for the high-security sample.

When we add up the deviations between the groups and those within them, as is done in Tables 12.2 and 12.3, we find that both are 0. This does not mean that there is an absence of variability either within or between the samples we are examining. Rather, this outcome reflects a rule stated in Chapter 4: The sum of the deviations from a mean equals 0. Clearly, we cannot use the sum of the deviations from the mean as an indicator of variation. As in other similar problems, it makes sense to square the deviations from the mean. The squares of all the deviations from the mean will be positive or 0.

The result when we square these quantities and then add them is commonly referred to as a **sum of squares.** The variability between groups measured in this way is called the **between sum of squares (BSS).** It is calculated by taking the sum of the squared deviation of each sample mean ($\overline{X}_c$) from the overall mean ($\overline{X}_g$) multiplied by the number of cases (N_c) in that sample.

$$\text{BSS} = \sum_{c=1}^{k} [N_c(\overline{X}_c - \overline{X}_g)^2]$$ **Equation 12.3**

Table 12.2 Summing the Deviations of the Group Means from the Grand Mean for the Three Groups in the Inmate Depression Study

$\overline{X}_c$	$\overline{X}_g$	$(\overline{X}_c - \overline{X}_g)$	$N_c(\overline{X}_c - \overline{X}_g)$
4	7	−3	−12
8	7	1	4
9	7	2	8
		$\sum = 0$	$\sum = 0$

Table 12.3 Summing the Deviations of the Scores from the Group Means Within the Three Groups in the Inmate Depression Study

X_i	$\overline{X}_c$	$(X_i - \overline{X}_c)$
3	4	−1
5	4	1
4	4	0
4	4	0
9	8	1
9	8	1
8	8	0
6	8	−2
9	9	0
10	9	1
7	9	−2
10	9	1
		$\sum = 0$

To calculate the between sum of squares for our hypothetical study, we take the same approach as is shown in Table 12.2. The one difference is that after we subtract the overall mean from a category mean, we square the result. Our final result is 56.

Working It Out

$$\text{BSS} = \sum_{c=1}^{k}[N_c(\overline{X}_c - \overline{X}_g)^2]$$

$$= 4(4 - 7)^2 + 4(8 - 7)^2 + 4(9 - 7)^2$$

$$= 56$$

When we measure variability within groups using this method, the result is defined as the **within sum of squares (WSS).** The within sum of squares is obtained by taking the sum of the squared deviation of each score from its category mean, as represented in Equation 12.4. As before, we first take the score for each subject and subtract from it the sample mean. However, before adding these deviations together, we square each one. The within sum of squares for our hypothetical example is equal to 14.

$$\text{WSS} = \sum_{i=1}^{N}(X_i - \overline{X}_c)^2 \qquad \textbf{Equation 12.4}$$

Working It Out

$$\begin{aligned}\text{WSS} &= \sum_{i=1}^{N} (X_i - \overline{X}_c)^2 \\ &= (3-4)^2 + (5-4)^2 + (4-4)^2 + (4-4)^2 + (9-8)^2 \\ &\quad + (9-8)^2 + (8-8)^2 + (6-8)^2 + (9-9)^2 + (10-9)^2 \\ &\quad + (7-9)^2 + (10-9)^2 \\ &= 14\end{aligned}$$

Partitioning Sums of Squares

We can also calculate a third type of variability for our inmate depression study: **total sum of squares (TSS).** The total sum of squares takes into account all of the variability in our three samples. It is calculated by taking the sum of the squared deviation of each score from the overall mean of scores for the three groups, as shown in Equation 12.5.

$$\text{TSS} = \sum_{i=1}^{N} (X_i - \overline{X}_g)^2 \qquad \textbf{Equation 12.5}$$

In practice, we first take the deviation of a score from the overall mean and then square it. For example, the first subject in the low-security prison sample has a score of 3. We subtract from this score the overall mean of 7 and then square the result (−4), to obtain a value of 16. To arrive at the total sum of squares, we do this for each of the 12 scores in the study and then sum the results.

Working It Out

$$\begin{aligned}\text{TSS} &= \sum_{i=1}^{N} (X_i - \overline{X}_g)^2 \\ &= (3-7)^2 + (5-7)^2 + (4-7)^2 + (4-7)^2 \\ &\quad + (9-7)^2 + (9-7)^2 + (8-7)^2 + (6-7)^2 + (9-7)^2 \\ &\quad + (10-7)^2 + (7-7)^2 + (10-7)^2 \\ &= 70\end{aligned}$$

The quantity obtained is equivalent to the sum of the between sum of squares and the within sum of squares. That is, the total variability across all of the scores is made up of the variability between the samples and the variability within the samples. More generally, the three types of variability discussed so far may be expressed in terms of a simple formula that partitions the total sum of squares into its two component parts: the between sum of squares and the within sum of squares.

$$\text{Total sum of squares} = \text{between sum of squares} + \text{within sum of squares}$$ **Equation 12.6**

(For our example, 70 = 56 + 14.)

Another way to express the relationship among the three types of sums of squares is to partition the total sum of squares into explained and unexplained components:

$$\text{Total sum of squares} = \text{explained sum of squares} + \text{unexplained sum of squares}$$ **Equation 12.7**

In this equation, the between sum of squares is represented by the **explained sum of squares (ESS)** because the between sum of squares represents the part of the total variability that is accounted for by the differences between the groups. For our hypothetical example, this is the proportion of the total variability in depression that is "explained" by the type of prison in which the subject is incarcerated.

The within sum of squares is represented in Equation 12.7 by the unexplained variability, or the **unexplained sum of squares (USS).** This is the part of the total variability that differences between the groups do not explain. We usually do not know the cause of this variability.

Developing Estimates of Population Variances

So far we have defined the sums of squares associated with between-group and within-group variability. But analysis of variance, as its name implies, is concerned with variance, not just variability. Accordingly, we have to adjust our sums by taking into account the appropriate number of degrees of freedom. In Chapter 5, when we developed estimates of variance, we divided the squared deviations from the mean by the number of cases in the sample or population. For analysis of variance, we divide the between and within sums of squares by the appropriate degrees of freedom.

For the between-group estimate of variance, we define the number of degrees of freedom as $k - 1$, where k is the number of samples or categories examined. If we are comparing three sample means, the number of degrees of freedom associated with the between-group estimate of variance is thus 2. As illustrated by Equation 12.8, an estimate of the

Representing Sums of Squares: ANOVA Notation in Different Formats

The ANOVA model is often presented using a different set of statistical notation than that used in this text. In this book, we define the total sum of squares as equal to the sum of the within sum of squares and the between sum of squares:

$$\sum_{i=1}^{N}(X_i - \overline{X}_g)^2 = \sum_{i=1}^{N}(X_i - \overline{X}_c)^2 + \sum_{c=1}^{k}[N_c(\overline{X}_c - \overline{X}_g)^2]$$

In many other statistics texts, the following equation is used for the decomposition of the total sum of squares:

$$\sum_{i=1}^{N}\sum_{j=1}^{k}(X_{ij} - \overline{X}_{\bullet\bullet})^2 = \sum_{i=1}^{N}\sum_{j=1}^{k}(X_{ij} - \overline{X}_{\bullet j})^2 + \sum_{i=1}^{N}\sum_{j=1}^{k}(\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet})^2$$

The double summation symbols indicate that we need to first sum over each group—denoted by the subscript j—and then sum over each individual observation—denoted by the subscript i. In terms of this book's notation, $\overline{X}_{\bullet\bullet} = \overline{X}_g$ and $\overline{X}_{\bullet j} = \overline{X}_c$.

Although we have reduced the number of summation signs to one for each sum of squares, the calculations with the two equations are identical for the total sum of squares and the within sum of squares. The one difference between the two equations lies in the calculation of the between sum of squares. The notation we offer simplifies this calculation by taking into account the fact that all of the individual scores in a single group (N_c) will have the same value. Rather than repeat the same calculation for all the individuals in the same group, we produce the identical answer by multiplying the squared difference of the sample mean and the overall mean by the number of observations in the corresponding sample.

between-group variance ($\hat{\sigma}^2_{bg}$) is obtained by dividing the between sum of squares by $k - 1$.

$$\hat{\sigma}^2_{bg} = \frac{\sum_{c=1}^{k}[N_c(\overline{X}_c - \overline{X}_g)^2]}{k - 1}$$ **Equation 12.8**

The number of degrees of freedom for the within-group estimate of variance is $N - k$, or the number of cases in the sample minus the number of samples or categories examined. The within-group variance estimate ($\hat{\sigma}^2_{wg}$) is calculated by dividing the within sum of squares by $N - k$ (see Equation 12.9).

$$\hat{\sigma}^2_{wg} = \frac{\sum_{i=1}^{N}(X_i - \overline{X}_c)^2}{N - k}$$ **Equation 12.9**

A Substantive Example: Age and White-Collar Crimes

Now that we have defined the two types of variance that are compared in analysis of variance, let's look at a substantive problem. Table 12.4 presents data on age for three samples of offenders convicted of white-collar crimes in seven federal district courts over a three-year period.[1]

Table 12.4 Ages of 30 White-Collar Criminals Convicted of Three Different Offenses

OFFENSE 1 BANK EMBEZZLEMENT	OFFENSE 2 BRIBERY	OFFENSE 3 ANTITRUST VIOLATION
19	28	35
21	29	46
23	32	48
25	40	53
29	42	58
30	48	61
31	58	62
35	58	62
42	64	62
49	68	75
$\overline{X} = 30.4$	$\overline{X} = 46.7$	$\overline{X} = 56.2$
$s = 8.98$	$s = 13.99$	$s = 10.54$
Grand mean ($\overline{X}_g$) = 1333/30 = 44.43		

[1]The data are drawn from S. Wheeler, D. Weisburd, and N. Bode, *Sanctioning of White Collar Crime, 1976–1978: Federal District Courts* (Ann Arbor, MI: Inter-University Consortium for Political and Social Research, 1988).

The first sample was drawn from offenders convicted of bank embezzlement, the second from offenders convicted of bribery, and the third from offenders convicted under criminal antitrust statutes. Ten subjects were drawn randomly from each of these populations.

The values listed in Table 12.4 represent the ages of the sampled offenders. The mean age of the bank embezzlers is 30.4 years; of the bribery offenders, 46.7 years; and of the antitrust offenders, 56.2 years. Can we conclude from the differences found among these samples that there are differences among the means of the populations from which these samples were drawn?

Assumptions:

Level of Measurement: Interval scale.

Population Distribution: Normal distribution for each population (the shape of the population distributions is unknown, and the sizes of the samples examined are small).

Sampling Method: Independent random sampling (no replacement; sample is small relative to population).

Sampling Frame: All white-collar offenders convicted of the crimes examined in seven federal judicial districts over a three-year period.

Population variances are equal ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2$).

Hypotheses:

H_0: Population means of age for bank embezzlers, bribery offenders, and antitrust offenders are equal ($\mu_1 = \mu_2 = \mu_3$).

H_1: Population means of age for bank embezzlers, bribery offenders, and antitrust offenders are not equal ($\mu_1 \neq \mu_2 \neq \mu_3$).

Like other parametric tests of statistical significance, analysis of variance requires that an interval level of measurement be used for the scores to be examined. Our example meets this assumption because age is an interval measure. Some statisticians add an assumption of nominal measurement because a comparison of means across samples requires that we define categories (or samples) for comparison. For example, in the case of age and white-collar crime, the three samples represent three categories of offenses. Our interest in this case is in the relationship between age (an interval-scale variable) and category of crime (a nominal-scale variable). In the hypothetical study discussed earlier in this chapter, we were interested in the relationship between depression (measured at an interval level) and type of prison (a nominal-scale variable).

Analysis of variance also requires that the populations underlying the samples examined be normally distributed. For our example, this is the most troubling assumption. We do not have evidence from prior studies that age is distributed normally within categories of crime. Nor are our samples large enough to allow us to invoke the central limit theorem. For ANOVA, as for the two-sample *t*-test, we want to have at least 30 cases per sample in order to relax the normality assumption. Because the computations for analysis of variance are complex, having only ten cases in each sample makes it easier to learn about ANOVA. However, our test will provide valid results only if the population distributions we infer to are in fact normally distributed.

We must also assume that the samples being compared were drawn randomly and independently. In practice, as we discussed in Chapter 11, researchers often make comparisons between groups within a single larger sample. For example, we might draw an independent random sample of white-collar offenders and then compare the means found in this larger sample for bank embezzlement, bribery, and antitrust offenders. As explained in Chapter 11, if the larger sample was drawn as an independent random sample, then we may assume that subsamples consisting of offenders convicted of different types of crimes are also independent random samples. In this study, random samples were drawn independently from each category of crime. Although the researchers did not sample with replacement, we can assume that this violation of assumptions is not serious because the sample drawn is very small relative to the population of offenders in the districts studied.

For analysis of variance, we must also assume that the population variances of the three groups are equal. This assumption of homoscedasticity is similar to that introduced for the two-sample *t*-test (using the pooled-variance method) in Chapter 11. However, in contrast to the *t*-test, ANOVA has no alternative test if we cannot assume equal variances between the groups. Although this seems at first to be an important barrier to using analysis of variance, in practice it is generally accepted that violations of this assumption must be very large before the results of a test come into question.[2]

One reason researchers, as opposed to statisticians, are not very concerned about the assumption of homoscedasticity is that even large

[2]For example, see G. Hornsnell, "The Effect of Unequal Group Variances on the *F* Test for Homogeneity of Group Means," *Biometrika* 40 (1954): 128–136; G. E. P. Box, "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems. I. Effect of Inequality of Variance in the One Way Classification," *Annals of Mathematical Statistics* 25 (1954): 290–302.

violations of this assumption affect the estimates of statistical significance to only a small degree. Sometimes it is suggested that you simply define a more conservative level of statistical significance when you are concerned with a serious violation of the homoscedasticity assumption.[3] Accordingly, you might select a 1% significance threshold as opposed to the more conventional 5% threshold. In general, large deviations in variance are not likely to occur across all of the groups studied. If one group is very different from the others, you might choose to conduct your test two ways—both including the group that is very different from the others and excluding it—and compare your results. In our example, the variances do not differ widely one from another (see Table 12.4).

Our final assumptions for this analysis relate to the null and research hypotheses. For analysis of variance, the null hypothesis is that the means of the groups are the same. In our example, the null hypothesis is that the mean ages of offenders in the populations of the three crime categories are equal. Our research hypothesis is that the means are not equal. As is true for ANOVA more generally, our research hypothesis is nondirectional. If we are making inferences to three or more populations, it is not possible to define the direction of the differences among them.[4]

The Sampling Distribution The sampling distribution used for making decisions about hypotheses in analysis of variance is called the *F* distribution, after R. A. Fisher, the statistician who first described it. The shape of the *F* distribution varies, depending on the number of degrees of freedom of the variance estimates being compared. The number of degrees

[3]Most packaged statistical programs provide a test for equivalence of variances as an option with their ANOVA program. However, be careful not to automatically reject use of analysis of variance on the basis of a statistically significant result. In smaller studies, with samples of less than 50 per group, a finding of a statistically significant difference should make you cautious about using analysis of variance. In such a study, you may want to adjust the significance level, as suggested here, or consider alternative nonparametric tests (discussed later in the chapter). With larger samples, a statistically significant result at conventional significance levels should not necessarily lead to any adjustments in your test. For such adjustments to be made, the difference should be highly significant and reflect large actual differences among variance estimates.

[4]However, in the special case of analysis of variance with only two samples, the researcher can use a directional research hypothesis. This will sometimes be done in experimental studies when the researcher seeks to examine differences across experimental and control groups, taking into account additional factors [e.g., see L. W. Sherman and D. Weisburd, “General Deterrent Effects of Police Patrol in Crime ‘Hot Spots.’ A Randomized Study,” *Justice Quarterly* 12 (1995): 625–648.]

of freedom is represented by $k - 1$ for the between-group variance and $N - k$ for the within-group variance:

Working It Out

$$\text{df for between-group variance} = k - 1 = 3 - 1 = 2$$

$$\text{df for within-group variance} = N - k = 30 - 3 = 27$$

Because we need to take into account two separate degrees of freedom at the same time, a different table of probability estimates is given for each significance threshold. Accordingly, Appendix 5 provides *F* tables for 0.05, 0.01, and 0.001 significance levels.

Each table provides the *F*-scores, adjusted for degrees of freedom, that correspond to the particular significance threshold. Thus, for example, in the table for $\alpha = 0.05$, the values given are the critical values for the test. For all tests using the *F* distribution, we need to obtain an *F*-score greater than this critical value to reject the null hypothesis of equal means. Looking at the table for $\alpha = 0.05$, we can identify two interesting characteristics of the *F* distribution.

First, the *F* distribution is unidirectional, consisting only of positive values. Consistent with the fact that the research hypothesis in analysis of variance with three or more population means states simply that the means are not equal, the *F* distribution is concerned only with the absolute size of the statistic obtained.

Second, as the number of degrees of freedom associated with the within-group variance grows, the *F*-value needed to reject the null hypothesis gets smaller. Remember that the number of degrees of freedom for the within-group variance is equal to $N - k$. Accordingly, as the number of cases in the sample gets larger, the number of degrees of freedom also gets larger. Why should the *F*-value needed to reject the null hypothesis be related to the size of the sample? As with a *t*-test, as the number of cases increases, so too does our confidence in the estimate we obtain from an *F*-test.[5]

Significance Level and Rejection Region Given that no special concerns have been stated in regard to the risk of either a Type I or a Type II error, we use a conventional 0.05 significance level. Looking at the *F* table for $\alpha = 0.05$ (Appendix 5), with 2 and 27 degrees of freedom,

[5]Indeed, note that the values of *F* with 1 degree of freedom for the between sum of squares are simply the values of *t* squared.

respectively, we find a critical value of 3.35. This tells us that we need an F-score greater than 3.35 to reject our null hypothesis of no difference between the population means. An observed F-score greater than 3.35 means that the observed significance level for our test is less than the 0.05 criterion level we have set.

The Test Statistic To calculate the F-ratio, we must compute estimates of the between-group and within-group population variances based on our three samples. Computing the between-group variance is relatively easy. As noted earlier, the formula for between-group variance ($\hat{\sigma}^2_{\text{bg}}$) is

$$\hat{\sigma}^2_{\text{bg}} = \frac{\sum_{c=1}^{k}[N_c(\overline{X}_c - \overline{X}_g)^2]}{k - 1}$$

Applying this formula to our example, we first take the mean for each group, subtract from it the overall mean of the three groups (44.43), square the result, and multiply by 10—the number of observations in each group. After this process has been carried out for each of the three groups, the totals are then added together and divided by the degrees of freedom for the between-group variance (2). These calculations are illustrated below. The between sum of squares for our example is 3,405.267. Dividing it by the number of degrees of freedom (2) results in a between-group variance estimate of 1,702.634.

Working It Out

$$\hat{\sigma}^2_{\text{bg}} = \frac{\sum_{c=1}^{k}[N_c(\overline{X}_c - \overline{X}_g)^2]}{k - 1}$$

$$= \frac{10(30.4 - 44.43)^2 + 10(46.7 - 44.43)^2 + 10(56.2 - 44.43)^2}{3 - 1}$$

$$= \frac{3{,}405.267}{2}$$

$$= 1{,}702.6335$$

Applying the formula for within-group variance is more difficult in large part because the calculation of a within-group sum of squares demands a good deal of computation even for small samples. For that reason, some texts provide an alternative estimating technique for the within-group sum of squares (see the box on p. 322). However, because it is probably safe to assume that you will turn to statistical computing packages when conducting research in the future and the purpose here is to gain a better understanding of analysis of variance, we will focus on the raw computation. Although cumbersome, it illustrates more directly the logic behind ANOVA.

As discussed earlier, the formula for within-group variance ($\hat{\sigma}^2_{wg}$) is

$$\hat{\sigma}^2_{wg} = \frac{\sum_{i=1}^{N}(X_i - \overline{X}_c)^2}{N - k}$$

For our example, we first take each individual score as illustrated in Table 12.5, and subtract from it the mean for its group: $(X_i - \overline{X}_c)$. We then square this quantity: $(X_i - \overline{X}_c)^2$. This is done for all 30 individual scores, and the results are then summed. The within-group sum of squares is 3,874.1. When we divide this quantity by the correct degrees of freedom ($N - k$, or 27), we obtain a within-group variance estimate of 143.485.

To obtain the F-statistic for our example, we simply calculate the ratio of the between- and within-group variances (see Equation 12.10), obtaining 11.866.

$$F = \frac{\text{between-group variance}}{\text{within-group variance}}$$ **Equation 12.10**

Working It Out

$$F = \frac{\text{between-group variance}}{\text{within-group variance}}$$

$$F = \frac{1{,}702.6335}{143.4852}$$

$$= 11.8663$$

Table 12.5 Calculating the Within-Group Sum of Squares

OFFENSE 1 BANK EMBEZZLEMENT $\overline{X} = 30.4$			OFFENSE 2 BRIBERY $\overline{X} = 46.7$			OFFENSE 3 ANTITRUST VIOLATION $\overline{X} = 56.2$		
X	$(X_i - \overline{X}_c)$	$(X_i - \overline{X}_c)^2$	X	$(X_i - \overline{X}_c)$	$(X_i - \overline{X}_c)^2$	X	$(X_i - \overline{X}_c)$	$(X_i - \overline{X}_c)^2$
19	−11.4	129.96	28	−18.7	349.69	35	−21.2	449.44
21	−9.4	88.36	29	−17.7	313.29	46	−10.2	104.04
23	−7.4	54.76	32	−14.7	216.09	48	−8.2	67.24
25	−5.4	29.16	40	−6.7	44.89	53	−3.2	10.24
29	−1.4	1.96	42	−4.7	22.09	58	1.8	3.24
30	−0.4	0.16	48	1.3	1.69	61	4.8	23.04
31	0.6	0.36	58	11.3	127.69	62	5.8	33.64
35	4.6	21.16	58	11.3	127.69	62	5.8	33.64
42	11.6	134.56	64	17.3	299.29	62	5.8	33.64
49	18.6	345.96	68	21.3	453.69	75	18.8	353.44
								$\sum = 3{,}874.1$

Computational Formulas for the Within-Group Sum of Squares

While it is important for you to understand the concepts underlying the equations used in the computations for ANOVA, the actual calculation of the within-group sum of squares can be quite tedious. Since it is often easier to calculate the total sum of squares and the between-group sum of squares, the simplest way of obtaining the within-group sum of squares is to rely on the relationship stated in Equation 12.6:

Total sum of squares = between-group sum of squares
+ within-group sum of squares

This equation can be rearranged and solved for the within-group sum of squares:

Within-group sum of squares = total sum of squares
− between-group sum of squares

A formula for computing the total sum of squares is

$$\text{TSS} = \sum_{i=1}^{N} X_i^2 - \frac{(\sum_{i=1}^{N} X_i)^2}{N}$$

This equation tells us to square the value of each observation and add the resulting squared values together $\left(\sum_{i=1}^{N} X_i^2\right)$. From this quantity, we then subtract the square of the sum of all the values divided by the total number of observations $\left((\sum_{i=1}^{N} X_i)^2/N\right)$.

For an illustration of the use of this formula, we can turn to the data on the ages of white-collar criminals in Table 12.4. In the following table, we take each offender's age (X) and square it (X^2). We then sum each column.

To calculate the total sum of squares, we just enter the two sums into the computational formula:

$$\begin{aligned}\text{TSS} &= \sum_{i=1}^{N} X_i^2 - \frac{(\sum_{i=1}^{N} X_i)^2}{N}\\ &= 66{,}509 - \frac{(1{,}333)^2}{30}\\ &= 7{,}279.367\end{aligned}$$

Age (X)	Age Squared (X^2)
19	361
21	441
23	529
25	625
29	841
30	900
31	961
35	1,225
42	1,764
49	2,401
28	784
29	841
32	1,024
40	1,600
42	1,764
48	2,304
58	3,364
58	3,364
64	4,096
68	4,624
35	1,225
46	2,116
48	2,304
53	2,809
58	3,364
61	3,721
62	3,844
62	3,844
62	3,844
75	5,625
$\sum = 1{,}333$	$\sum = 66{,}509$

At this point, since we know the total sum of squares (7,279.367) and have already calculated the between-group sum of squares (3,405.267), we can easily see that the within-group sum of squares is 3,874.1:

$$
\begin{aligned}
\text{Within-group sum of squares} &= \text{total sum of squares} \\
&\quad - \text{between-group sum of squares} \\
&= 7{,}279.367 - 3{,}405.267 \\
&= 3{,}874.1
\end{aligned}
$$

The Decision Because the test statistic for our example (11.866) is larger than 3.35, the critical value of the rejection region, our result is statistically significant at the 0.05 level. Accordingly, we reject the null hypothesis of no difference between the population means and conclude (with a conventional level of risk of falsely rejecting the null hypothesis) that the average age of offenders differs across the three types of crime examined. However, given our concern about violating the assumption of normality, our conclusion will be valid only if age is indeed normally distributed in the three populations studied.

Another ANOVA Example: Race and Bail Amounts Among Felony Drug Defendants

Table 12.6 presents data on bail amounts set for three samples of felony drug defendants in large urban court districts in the United States in the 1990s.[6] The first sample is taken from non-Hispanic whites, the second sample from non-Hispanic African Americans, and the third sample from Hispanics of any race. Twelve defendants were drawn at random from the population of each group. The mean bail amount is \$4,833.33 for non-Hispanic whites, \$8,833.33 for non-Hispanic African Americans, and \$30,375.00 for Hispanics of any race. Do

Table 12.6 Bail Amounts (in Dollars) for 36 Felony Drug Defendants

NON-HISPANIC WHITES	NON-HISPANIC BLACKS	HISPANICS OF ANY RACE
1,000	1,000	1,000
1,000	1,000	2,000
1,500	2,000	4,000
2,000	2,500	5,000
2,500	3,000	10,000
3,000	4,000	12,500
5,000	5,000	25,000
7,000	10,000	25,000
7,500	12,500	25,000
7,500	20,000	40,000
10,000	20,000	65,000
10,000	25,000	150,000
$\overline{X} = 4{,}833.33$	$\overline{X} = 8{,}833.33$	$\overline{X} = 30{,}375.00$
$s = 3{,}287.18$	$s = 8{,}201.46$	$s = 42{,}028.74$
Grand mean ($\overline{X}_g$) = 528500/36 = 14,680.56		

[6]The data are taken from *State Court Processing Statistics: 1990, 1992, 1994, 1996* and can be accessed through the National Archive of Criminal Justice Data web site at http://www.icpsr.umich.edu/NACJD.

The ANOVA Table

Most statistical analysis software presents the results of an analysis of variance in the form of an ANOVA table. An ANOVA table provides a compact and convenient way to present the key elements in an analysis of variance. In addition to ensuring that the researcher has all the necessary information, it also allows the researcher to reproduce the estimates of the variance and the *F*-statistic. The general form of an ANOVA table is as follows:

Source	**df**	**Sum of Squares**	**Mean Square**	***F***
Between	$k - 1$	$\sum_{c=1}^{k}[N_c(\bar{X}_c - \bar{X}_g)^2]$	$\frac{\text{BSS}}{k-1}$	$\frac{\hat{\sigma}^2_{bg}}{\hat{\sigma}^2_{wg}}$
Within	$N - k$	$\sum_{i=1}^{N}(X_i - \bar{X}_c)^2$	$\frac{\text{WSS}}{N-k}$	
Total	$N - 1$	$\sum_{i=1}^{N}(X_i - \bar{X}_g)^2$		

Each row gives the pieces of information needed and the formulas for the calculations. For example, the "Between" row gives the corresponding degrees of freedom and the formulas for calculating between-group variability, between-group variance (in the "Mean Square" column), and the *F*-statistic. The "Within" row gives the corresponding degrees of freedom and the formulas for calculating within-group variability and within-group variance.

Following is the ANOVA table for the results of our calculations using the data on the ages of white-collar criminals:

Source	**df**	**Sum of Squares**	**Mean Square**	***F***
Between	2	3,405.267	1,702.6335	11.8663
Within	27	3,874.100	143.4852	
Total	29	7,279.367		

the differences in sample means indicate that there are differences in the population means?

Assumptions:

Level of Measurement: Interval scale.

Population Distribution: Normal distribution for each population (the shape of the population distributions is unknown, and the sizes of the samples examined are small).

Sampling Method: Independent random sampling (no replacement; samples are small relative to populations).

Sampling Frame: Felony drug defendants in large urban court districts in the United States in the 1990s.

Population variances are equal ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2$).

Hypotheses:

H_0: Population means of bail amounts set for felony drug defendants who are non-Hispanic whites, non-Hispanic African Americans, and Hispanics of any race are equal ($\mu_1 = \mu_2 = \mu_3$).

H_1: Population means of bail amounts set for felony drug defendants who are non-Hispanic whites, non-Hispanic African Americans, and Hispanics of any race are not equal ($\mu_1 \neq \mu_2 \neq \mu_3$).

The Sampling Distribution We again use the F distribution to test for differences among our three sample means. Recall that we need two indicators of degrees of freedom: one for the between-group variance and one for the within-group variance.

$$\text{df for between-group variance} = k - 1 = 3 - 1 = 2$$

$$\text{df for within-group variance} = N - k = 36 - 3 = 33$$

Significance Level and Rejection Region As in the preceding example, we do not have any particular concerns about the risk of a Type I or Type II error, so we can use the conventional 0.05 significance level. Given that we have degrees of freedom equal to 2 and 33 with a 0.05 significance level, the critical value of the F-statistic is about 3.29. If our calculated F-statistic is greater than 3.29, then we will reject our null hypothesis of equal means.

The Test Statistic We begin our calculation of the F-statistic by computing estimates of the between-group variance and the within-group variance. Applying the formula for between-group variance, we find the estimate of $\hat{\sigma}^2_{\text{bg}}$ to be 2,264,840,905.56.

Working It Out

$$\hat{\sigma}^2_{\text{bg}} = \frac{\sum_{c=1}^{k}[N_c(\overline{X}_c - \overline{X}_g)^2]}{k - 1}$$

$$= \frac{\left(\begin{array}{c}12(4{,}833.33 - 14{,}680.56)^2 \\ + 12(8{,}833.33 - 14{,}680.56)^2 + 12(30{,}375.00 - 14{,}680.56)^2\end{array}\right)}{3 - 1}$$

$$= \frac{4{,}529{,}681{,}811.11}{2}$$

$$= 2{,}264{,}840{,}905.56$$

The value of the within-group variance ($\hat{\sigma}^2_{\text{wg}}$) is 617,193,813.13. Table 12.7 presents the calculation of the within-group sum of squares, which turns out to be equal to 20,367,395,833.33. We then divide the value of the within-group sum of squares by the corresponding degrees of freedom ($N - k = 36 - 3 = 33$), which gives us an estimate for the within-group variance of 617,193,813.13.

Working It Out

$$\hat{\sigma}^2_{\text{wg}} = \frac{\sum_{i=1}^{N}(X_i - \overline{X}_c)^2}{N - k}$$

$$= \frac{20{,}367{,}395{,}833.33}{36 - 3}$$

$$= 617{,}193{,}813.13$$

The value of the F-statistic is obtained by dividing the estimate of between-group variance by the estimate of within-group variance. For our example, F is found to be 3.67.

Working It Out

$$F = \frac{\text{between-group variance}}{\text{within-group variance}}$$

$$= \frac{2{,}264{,}840{,}905.56}{617{,}193{,}813.13}$$

$$= 3.67$$

Table 12.7 Calculating the Within-Group Sum of Squares

NON-HISPANIC WHITES		NON-HISPANIC BLACKS		HISPANICS OF ANY RACE	
X	$(X_i - \bar{X}_c)^2$	X	$(X_i - \bar{X}_c)^2$	X	$(X_i - \bar{X}_c)^2$
1,000	14,694,418.89	1,000	61,361,058.89	1,000	862,890,625.00
1,000	14,694,418.89	1,000	61,361,058.89	2,000	805,140,625.00
1,500	11,111,088.89	2,000	46,694,398.89	4,000	695,640,625.00
2,000	8,027,758.89	2,500	40,111,068.89	5,000	643,890,625.00
2,500	5,444,428.89	3,000	34,027,738.89	10,000	415,140,625.00
3,000	3,361,098.89	4,000	23,361,078.89	12,500	319,515,625.00
5,000	27,778.89	5,000	14,694,418.89	25,000	28,890,625.00
7,000	4,694,458.89	10,000	1,361,118.89	25,000	28,890,625.00
7,500	7,111,128.89	12,500	13,444,468.89	25,000	28,890,625.00
7,500	7,111,128.89	20,000	124,694,518.89	40,000	92,640,625.00
10,000	26,694,478.89	20,000	124,694,518.89	65,000	1,198,890,625.00
10,000	26,694,478.89	25,000	261,361,218.89	150,000	14,310,140,625.00
				$\sum$ = 20,367,395,833.33	

The Decision Since our test statistic of 3.67 is greater than the critical value of 3.29 for the F distribution, we reject the null hypothesis of equal population means; we conclude that there is a statistically significant relationship between bail amount and race of defendant. However, as in our example concerning age and white-collar crime, we began our test with strong doubts about whether we could meet a core assumption of analysis of variance. The samples examined are not large, and thus we cannot relax the normality assumption for our test. In turn, we do not have strong reason to believe that the populations to which we want to infer actually meet the criteria of this assumption. As has been noted throughout the text, statistical conclusions are only as solid as the assumptions the researchers make. In this case, our statistical conclusions clearly do not stand on solid ground.

Defining the Strength of the Relationship Observed

Even though analysis of variance is concerned with comparing means from independent samples, in practice the samples are usually defined as representing a multicategory nominal-level variable. For example, as noted earlier, in comparing three samples of white-collar criminals, we could define each as one category in a nominal-scale measure of type of white-collar crime. Similarly, in our example concerning the relationship between bail amount and race, we spoke about differences among three samples of offenders: non-Hispanic whites, non-Hispanic blacks, and

Hispanics of any race. Nonetheless, these three samples can be seen as three groups in a nominal-level measure of race of defendant.

Accordingly, one question we might ask after finding a statistically significant result is "How strong is the overall relationship we have identified?" The simplest way to answer this question is to look at the differences between the means of the samples. With just three samples, we can get a pretty good sense of the strength of a relationship using this method. But even with three samples, it is difficult to summarize the extent of the relationship observed because we must look at three separate comparisons (that between group 1 and group 2, that between group 2 and group 3, and that between group 3 and group 1). With four samples, the number of comparisons is six; for seven samples, there are 21 comparisons. Clearly, it is useful, especially as the number of samples grows, to have a single statistic for establishing the strength of the observed relationship.

A commonly used measure of association for ANOVA is a statistic called **eta (η).** Eta relies on the partialing of sums of squares to establish the relationship, or **correlation,** between the interval-level variable in ANOVA and the nominal-level variable. To calculate eta, we simply take the square root of the ratio of the between sum of squares to the total sum of squares (see Equation 12.11).

$$\eta = \sqrt{\frac{\text{BSS}}{\text{TSS}}}$$

Equation 12.11

Although it might not seem so at first glance, this measure makes good sense. Understanding eta, however, will be easier if we start with another statistic, **eta squared (η^2),** which is sometimes referred to as the **percent of variance explained** (see Equation 12.12).

$$\eta^2 = \frac{\text{BSS}}{\text{TSS}}$$

Equation 12.12

Eta squared is the proportion of the total sum of squares that is accounted for by the between sum of squares. As previously noted, the between sum of squares is also defined as the explained sum of squares because it represents the part of the total variation that is accounted for by the differences between the samples. Eta squared thus identifies the proportion of the total sum of squares that is accounted for by the explained sum of squares—hence its identification as the percent of variance explained.

The larger the proportion of total variance that is accounted for by the between sum of squares, the stronger the relationship between the nominal- and interval-level variables being examined. When the means of the samples are the same, eta squared will be 0. This means that there is no relationship between the nominal- and interval-level measures the study is examining. The largest value of eta squared is 1, meaning that all of

Table 12.8 Comparing Eta Squared with Eta

η^2	η
0.00	0.00
0.01	0.10
0.02	0.14
0.03	0.17
0.04	0.20
0.05	0.22
0.10	0.32
0.25	0.50
0.50	0.71
0.75	0.87
1.00	1.00

the variability in the samples is accounted for by the between sum of squares. In practice, as eta squared increases in value between 0 and 1, the relationship being examined gets stronger.

The square root of eta squared is a measure more sensitive to small relationships. For example, a value for eta squared of 0.04 is equal to a value for eta of 0.20, and a value for eta squared of 0.1 is equivalent to a value for eta of 0.32, as shown in Table 12.8. In criminal justice, where the relationships examined are often not very large, measures such as this one, which allow us to distinguish relatively small values more clearly, are particularly useful.

Turning to our example concerning age and white-collar crime, we can see that the differences between the groups account for a good deal of the variability in the total sum of squares. Taking the between sum of squares for that example and dividing it by the total sum of squares gives a value for eta squared of 0.468.

Working It Out

$$\text{BSS} = 3{,}405.267$$

$$\begin{aligned}\text{TSS} &= \text{BSS} + \text{WSS} \\ &= 3{,}405.267 + 3{,}874.100 \\ &= 7{,}279.367\end{aligned}$$

$$\begin{aligned}\eta^2 &= \frac{\text{BSS}}{\text{TSS}} \\ &= \frac{3{,}405.267}{7{,}279.367} \\ &= 0.4678\end{aligned}$$

By taking the square root of this value, we obtain a correlation coefficient, or eta, of 0.684.

Working It Out

$$\eta = \sqrt{\frac{\text{BSS}}{\text{TSS}}} = \sqrt{\frac{3{,}405.267}{7{,}279.367}}$$

$$= \sqrt{0.4678}$$

$$= 0.6840$$

Does an eta of 0.684 signify a large or small relationship? To some extent, differentiating between "large" and "small" in this context is a value judgment rather than a statistical decision. We might decide whether a particular value of eta is large or small based on results from other studies in other areas of criminal justice or perhaps similar studies that drew different samples. There is no clear yardstick for making this decision. One psychologist suggests that any value for eta greater than 0.371 represents a large effect.[7] A moderate-size effect is indicated by a value of 0.243. Using this criterion, we would define the relationship between age and type of white-collar crime as very strong. However, in this example, we should be cautious about relying on the results obtained. With small samples, the values of eta are not considered very reliable.[8]

Making Pairwise Comparisons Between the Groups Studied

Once you have established through an analysis of variance that there is a statistically significant difference across the samples studied, you may want to look at differences between specific pairs of the samples. To do this, you make comparisons between two sample means at a time. Such comparisons within an analysis of variance are often called **pairwise comparisons.**

[7]See Jacob Cohen, *Statistical Power Analysis for the Behavorial Sciences* (Hillsdale, NJ: Lawrence Erlbaum, 1988), pp. 285–287.

[8]Once again, there is no universally accepted definition of what is "small." There will be little question regarding the validity of your estimate of eta if your samples meet the 30 cases minimum defined for invoking the central limit theorem. Some statisticians suggest that you will gain relatively reliable estimates even for samples as small as 10 cases.

Table 12.9 The 21 Separate Pairwise Comparisons To Be Made for an Analysis of Variance with Seven Samples (Categories)

SAMPLE	1	2	3	4	5	6	7
1							
2	√						
3	√	√					
4	√	√	√				
5	√	√	√	√			
6	√	√	√	√	√		
7	√	√	√	√	√	√	

It would seem, at first glance, that you could simply apply the two-sample *t*-test discussed in Chapter 11 to test hypotheses related to such comparisons. However, you are faced with a very important statistical problem. If you run a number of *t*-tests at the same time, you are unfairly increasing your odds of obtaining a statistically significant finding along the way. For example, let's say that you have conducted an analysis of variance comparing seven samples and obtained a statistically significant result. You now want to look at the pairwise comparisons to see which of the specific comparisons are significantly different from one another. There are a total of 21 separate comparisons to make (see Table 12.9). For each test, you set a significance level of 0.05, which means that you are willing to take a 1 in 20 chance of falsely rejecting the null hypothesis. Thus, if you run 20 tests, you might expect to get at least one statistically significant result just by chance.

Here a finding of a significant result could simply be attributed to the fact that you have run a large number of tests. Accordingly, to be fair, you should adjust your tests to take into account the change in the probabilities that results from looking at a series of pairwise comparisons. A number of different tests allow you to do this; many are provided in standard statistical packages.[9] One commonly used test is the **honestly significant difference (HSD) test** developed by John Tukey (see Equation 12.13).

$$\text{HSD} = P_{\text{crit}} \sqrt{\frac{\hat{\sigma}^2_{\text{wg}}}{N_c}}$$ **Equation 12.13**

HSD defines the value of the difference between the pairwise comparisons that is required to reject the null hypothesis at a given level of statistical significance.

[9]For a discussion of pairwise comparison tests, see A. J. Klockars and G. Sax, *Multiple Comparisons* (*Quantitative Applications in the Social Science,* Vol. 61) (London: Sage, 1986).

For our white-collar crime example, with a conventional 5% significance threshold, we first identify the critical value (P_{crit}) associated with that significance threshold by looking at Appendix 6. With three samples and 27 degrees of freedom in the within sum of squares estimate, the critical value is about 3.51. We then multiply this value by the square root of the within-group variance estimate ($\hat{\sigma}^2_{\text{wg}}$) divided by the number of cases in each sample (N_c)—in our example, 10.[10] Our result is 13.296, meaning that the absolute value of the difference in mean age between the pairwise comparisons must be greater than 13.296 to reject the null hypothesis of no difference (using a 5% significance threshold).

Working It Out

$$\text{HSD} = P_{\text{crit}} \sqrt{\frac{\hat{\sigma}^2_{\text{wg}}}{N_c}}$$

$$= 3.51 \sqrt{\frac{143.49}{10}}$$

$$= 13.296$$

Table 12.10 shows the absolute differences found for the three comparisons between means. Two of the three comparisons are statistically significant at the 5% level—the absolute differences are greater than 13.296 for the difference between bank embezzlers and bribery offenders and for that between bank embezzlers and antitrust offenders. However, for the difference between bribery and antitrust offenders, our result just misses the value needed to reject the null hypothesis. We would have to conclude that our sample results do not provide persuasive evidence for stating that the mean ages of bribery and antitrust offenders are different in the larger populations from which these two samples were drawn.

---

[10]Most pairwise comparison tests, including Tukey's HSD test, require that the sample sizes of the groups examined be equal. While most statistical software packages provide adjustments of these tests to account for unequal sample sizes, there is still debate over whether the estimates gained can be relied upon [e.g., see Robert R. J. Sokal and F. J. Rohlf, *Biometry: The Principles and Practice of Statistics in Biological Research*, 3rd ed. (New York: W. H. Freeman, 1995), Chap. 9]. Irrespective of this debate, when unequal sample sizes are examined, the adjusted estimates are to be preferred over the unadjusted estimates.

Table 12.10 Results of the Pairwise Comparison Tests

	OFFENSE 1 BANK EMBEZZLEMENT	OFFENSE 2 BRIBERY	OFFENSE 3 ANTITRUST VIOLATION
Offense 1 Bank Embezzlement			
Offense 2 Bribery	16.300*		
Offense 3 Antitrust Violation	25.800*	9.500	

*$p < 0.5$

The comparisons we have made so far have been based on a statistically significant overall result for ANOVA. Should such comparisons be made if the overall differences across the means are not statistically significant? In general, it is not a good idea to look for pairwise comparisons if the overall analysis of variance is not statistically significant. This is a bit like going fishing for a statistically significant result. However, sometimes one or another of the pairwise comparisons is of particular interest. Such interest should be determined before you develop your analyses. However, if you do start off with a strong hypothesis for a pairwise comparison, it is acceptable to examine it, irrespective of the outcomes of the larger test. In such circumstances, it is also acceptable to use a simple two-sample *t*-test to examine group differences.

A Nonparametric Alternative: The Kruskal-Wallis Test

For studies where you cannot meet the parametric assumptions of the analysis of variance test, you may want to consider a nonparametric **rank-order test.** In performing a rank-order test, you lose some crucial information because you focus only on the order of scores and not on the differences in values between them. However, such tests have the advantage of not requiring assumptions about the population distribution.

One rank-order test is the **Kruskal-Wallis test.** As a nonparametric test of statistical significance, it requires neither a normal distribution nor equal variances between the groups studied. The test asks simply whether the distribution of ranked scores in the three groups is what would be expected under a null hypothesis of no difference. When the number of cases in each group is greater than 5, the sampling distribution of the Kruskal-Wallis test score, denoted H, is approximately chi-square.

As an illustration, let's examine whether this nonparametric test suggests significant differences in terms of age across the white-collar crime categories of our earlier example.

Assumptions:

Level of Measurement: Ordinal scale.

Population Distribution: No assumption made.

Sampling Method: Independent random sampling (no replacement; sample is small relative to population).

Sampling Frame: All white-collar offenders convicted of the crimes examined in seven federal judicial districts over a three-year period.

Hypotheses:

H_0: The distribution of ranked scores is identical in the three populations.

H_1: The distribution of ranked scores differs across the three populations.

In this test, we use an ordinal-level measure: the rank order of ages in the sample. To obtain this measure, we simply rank the 30 subjects studied according to age, with the youngest offender given a rank of 1 and the oldest a rank of 30 (see Table 12.11). In the case of ties, subjects share a rank. For example, the two subjects aged 29 share the rank of 6.5 (the average of ranks 6 and 7), and the three subjects aged 62 share the rank of 26 (the average of ranks 25, 26, and 27).

Table 12.11 White-Collar Offenders Ranked According to Age

OFFENSE 1 BANK EMBEZZLEMENT		OFFENSE 2 BRIBERY		OFFENSE 3 ANTITRUST VIOLATION	
Age	**Rank**	**Age**	**Rank**	**Age**	**Rank**
19	1	28	5	35	11.5
21	2	29	6.5	46	16
23	3	32	10	48	17.5
25	4	40	13	53	20
29	6.5	42	14.5	58	22
30	8	48	17.5	61	24
31	9	58	22	62	26
35	11.5	58	22	62	26
42	14.5	64	28	62	26
49	19	68	29	75	30
	$\sum = 78.5$		$\sum = 167.5$		$\sum = 219$

Our null hypothesis is that the distribution of ranked scores in the three populations is identical. Our research hypothesis is that it is not identical across the three populations.

Sampling Distribution The sampling distribution *H* is distributed approximately according to chi-square because the number of cases in each group is greater than 5. The number of degrees of freedom for the distribution is defined as $k - 1$, where *k* refers to the number of samples (or categories). Because our example involves three samples, the number of degrees of freedom for the chi-square distribution is $3 - 1$, or 2.

Significance Level and Rejection Region Consistent with our earlier choice of a 0.05 significance threshold, we turn to the 0.05 value with 2 degrees of freedom in the chi-square table (see Appendix 2). The critical value identified is 5.991.

The Test Statistic The formula for *H* given in Equation 12.14 looks complex. However, it is relatively simple to compute if broken into pieces.

$$H = \left[\left(\frac{12}{N(N+1)} \right) \left(\sum_{c=1}^{k} \frac{\left(\sum_{i=1}^{N_c} R_i \right)^2}{N_c} \right) \right] - 3(N+1) \qquad \textbf{Equation 12.14}$$

There is only one complex term in the equation. It is

$$\sum_{c=1}^{k} \frac{\left(\sum_{i=1}^{N_c} R_i \right)^2}{N_c}$$

This term tells us to take the sum of the ranks in each sample, square it, and divide it by the number of cases in the sample; then we sum these values for all the samples.[11] The *H*-score obtained for our problem is 13.038.

[11] Most statistical computing packages provide an alternative calculation that adjusts for ties. In practice, the differences between using this correction procedure and performing the unadjusted test are generally small. For our example, where there are a large number of ties relative to the sample size (14/30), the difference in the observed significance level is only 0.0001.

Working It Out

$$H = \left[\left(\frac{12}{N(N+1)}\right)\left(\sum_{c=1}^{k}\frac{\left(\sum_{i=1}^{N_c} R_i\right)^2}{N_c}\right)\right] - 3(N+1)$$

$$= \left(\frac{12}{30(31)}\right)\left(\frac{(78.5)^2}{10} + \frac{(167.5)^2}{10} + \frac{(219)^2}{10}\right) - 3(31)$$

$$= \left(\frac{12}{930}\right)8{,}217.95 - 3(31)$$

$$= 13.038$$

The Decision As with the *F*-test, our score exceeds the critical value needed to reject the null hypothesis of no difference. The observed significance level of our test is thus less than the criterion significance level we set at the outset ($p < 0.05$). From the Kruskal-Wallis test, we can again conclude that there is a statistically significant relationship between type of white-collar crime and age of offender. This time, however, we can have more confidence in our conclusion because the assumptions of the test are met more strictly.

Chapter Summary

ANOVA is a parametric test of statistical significance that allows a researcher to compare means across more than two groups. It takes into account not only variability between groups but also variability within groups. The larger the differences between the groups relative to the variability within them, the more confidence the researcher can have in a conclusion that differences exist among the population means. Between-group variability is measured by the **between sum of squares** (or **explained sum of squares**). Within-group variability is measured by the **within sum of squares** (or **unexplained sum of squares**). The **total sum of squares** is equal to the sum of the between and within sums of squares. To develop estimates of population variances, the sums of squares are divided by the appropriate degrees of freedom. ANOVA requires the following assumptions: interval scales, normal population distributions, independent random sampling, and homoscedasticity. The sampling distribution for ANOVA is denoted as *F*.

The *F*-value needed to reject the null hypothesis gets smaller as the within-group degrees of freedom grows. The *F*-statistic is calculated by dividing the between-group variance by the within-group variance.

The strength of the relationship observed is measured by the statistic **eta squared,** or **percent of variance explained.** Eta squared is the ratio of the between sum of squares to the total sum of squares. An eta squared value of 0 indicates that there is no relationship between the nominal and interval variables (i.e., the means are the same). An eta squared value of 1 represents a perfect relationship between the nominal and interval variables. The **correlation coefficient,** or **eta,** is obtained by taking the square root of eta squared.

A researcher who wishes to compare means of pairs of specific samples within a larger test makes a **pairwise comparison.** Running a series of two-sample *t*-tests, however, unfairly increases the odds of getting a statistically significant result. The **honestly significant difference (HSD) test** is a pairwise comparison test that corrects for this bias.

When the assumptions underlying ANOVA are difficult to meet, the researcher may choose a nonparametric alternative—the **Kruskal-Wallis test.** This test does not require an assumption of normal population distributions or homoscedasticity. As a **rank-order test,** however, it does not use all of the information available from interval-level data.

Key Terms

analysis of variance (ANOVA) A parametric test of statistical significance that assesses whether differences in the means of several samples (groups) can lead the researcher to reject the null hypothesis that the means of the populations from which the samples are drawn are the same.

between sum of squares (BSS) A measure of the variability between samples (groups). The between sum of squares is calculated by taking the sum of the squared deviation of each sample mean from the grand mean multiplied by the number of cases in that sample.

correlation A measure of the strength of a relationship between two variables.

eta A measure of the degree of correlation between an interval-level and a nominal-level variable.

eta squared The proportion of the total sum of squares that is accounted for by the between sum of squares. Eta squared is sometimes referred to as the percent of variance explained.

explained sum of squares (ESS) Another name for the between sum of squares. The explained sum of squares is the part of the total variability that can be explained by visible differences between the groups.

grand mean The overall mean of every single case across all of the samples.

honestly significant difference (HSD) test A parametric test of statistical significance, adjusted for making pairwise comparisons. The HSD test defines the difference between the pairwise comparisons required to reject the null hypothesis.

Kruskal-Wallis test A nonparametric test of statistical significance for multiple groups, requiring at least an ordinal scale of measurement.

overall mean See *grand mean.*

pairwise comparisons Comparisons made between two sample means extracted from a larger statistical analysis.

percent of variance explained The proportion of the total sum of squares that is accounted for by the explained sum of squares; eta squared.

rank-order test A test of statistical significance that uses information relating to the relative order, or rank, of variable scores.

sum of squares The sum of squared deviations of scores from a mean or set of means.

total sum of squares (TSS) A measure of the total amount of variability across all of the groups examined. The total sum of squares is calculated by summing the squared deviation of each score from the grand mean.

unexplained sum of squares (USS) Another name for the within sum of squares. The unexplained sum of squares is the part of the total variability that cannot be explained by visible differences between the groups.

within sum of squares (WSS) A measure of the variability within samples (groups). The within sum of squares is calculated by summing the squared deviation of each score from its sample mean.

Symbols and Formulas

X_i	Individual subject or score
$\overline{X}_c$	Sample or category mean
$\overline{X}_g$	Grand or overall mean
N_c	Number of cases in each sample
k	Number of categories or samples
η	Correlation coefficient eta
η^2	Percent of variance explained; eta squared
P_{crit}	Critical value for HSD test
R_i	Individual rank of score

To calculate the between sum of squares:

$$\text{BSS} = \sum_{c=1}^{k} [N_c(\overline{X}_c - \overline{X}_g)^2]$$

To calculate the within sum of squares:

$$\text{WSS} = \sum_{i=1}^{N}(X_i - \overline{X}_c)^2$$

To calculate the total sum of squares:

$$\text{TSS} = \sum_{i=1}^{N}(X_i - \overline{X}_g)^2$$

To partition the total sum of squares:

Total sum of squares = between sum of squares + within sum of squares

$$\text{TSS} = \text{BSS} + \text{WSS}$$

To estimate between-group variance:

$$\hat{\sigma}^2_{\text{bg}} = \frac{\sum_{c=1}^{k}[N_c(\overline{X}_c - \overline{X}_g)^2}{k - 1}$$

To estimate within-group variance:

$$\hat{\sigma}^2_{\text{wg}} = \frac{\sum_{i=1}^{N}(X_i - \overline{X}_c)^2}{N - k}$$

To calculate F:

$$F = \frac{\text{between-group variance}}{\text{within-group variance}}$$

To calculate eta:

$$\eta = \sqrt{\frac{\text{BSS}}{\text{TSS}}}$$

To calculate eta squared:

$$\eta^2 = \frac{\text{BSS}}{\text{TSS}}$$

To perform the HSD test:

$$\text{HSD} = P_{\text{crit}}\sqrt{\frac{\hat{\sigma}^2_{\text{wg}}}{N_c}}$$

To perform the Kruskal-Wallis test:

$$H = \left[\left(\frac{12}{N(N+1)}\right)\left(\sum_{c=1}^{k}\frac{\left(\sum_{i=1}^{N_c} R_i\right)^2}{N_c}\right)\right] - 3(N+1)$$

Exercises

12.1 Dawn, a criminal justice researcher, gives 125 pretrial defendants scores based on a questionnaire that assesses their ability to understand the court process. The defendants were selected from five separate counties. Dawn took an independent random sample of 25 defendants from each county. The scores for the five populations are normally distributed. Dawn runs an ANOVA test on her results, which produces a test statistic of 3.35.

a. Would Dawn be able to reject her null hypothesis that there is no difference between the populations in their ability to comprehend the court process if she were to set a 5% significance level?

b. Would she be able to reject the null hypothesis using a 1% significance level?

c. Would your answer to either part a or part b be different if Dawn's sample had consisted of five equally sized groups of 200 subjects each?

12.2 Random samples of individuals were drawn from three neighborhoods by a policing research foundation to study the level of public support for the local police department. The research foundation constructed a complicated interval-level measure of police support, in which higher values indicated more support. The researchers found the following pattern across the three neighborhoods: The mean level of support in neighborhood A was 3.1 ($N = 15$); in neighborhood B, it was 5.6 ($N = 17$); and in neighborhood C, 4.2 ($N = 11$). The measure of between-group variance was 4.7, and the measure of within-group variance was 1.1.

a. If the significance level is 0.05, can the research foundation conclude that there are different levels of support for the police department across neighborhoods? Write out all of the steps of a test of statistical significance, including any violations of assumptions.

b. What if the significance level is 0.01?

12.3 Random sampling of individuals with four different majors at a university found the following grade point averages (GPAs) for the four groups:

Major A: GPA = 3.23 ($N = 178$)

Major B: GPA = 2.76 ($N = 64$)

Major C: GPA = 2.18 ($N = 99$)

Major D: GPA = 3.54 ($N = 121$)

If the between-group variance is 5.7 and the within-group variance is 1.5, are the GPAs different for the four majors? Use a significance level of 0.01, and write out all of the steps of a test of statistical significance, including any violations of assumptions.

12.4 Random sampling offenders convicted of minor drug possession in Border County found the average jail sentence for white offenders to be 86 days ($N = 15$), for African American offenders to be 99 days ($N = 10$), and for Hispanic offenders to be 72 days ($N = 7$). Further analysis of jail sentence lengths by race found the between sum of squares to be 250 and the within sum of squares to be 1,300. Are the jail sentence lengths significantly different across race categories?

a. Use a significance level of 0.05. Write out all of the steps of a test of statistical significance, including any violations of assumptions.

b. Would the conclusion be any different if the significance level had been set at 0.01?

12.5 Listed below is a set of data identifying previous convictions for any offense of 40 inmates serving prison sentences for robbery, rape, murder, and drug dealing.

Robbery	Rape	Murder	Drug Dealing
1	1	0	5
0	1	0	3
2	1	0	7
6	0	6	4
4	0	2	8
5	2	7	0
3	2	1	6
1	1	4	2
5	0	2	1
3	2	3	4

Calculate the following values:

a. $\overline{X}_g$

b. df for between-group variance

c. df for within-group variance

d. the four values of $\overline{X}_c$

e. the total sum of squares

f. the between sum of squares

g. the within sum of squares

12.6 Convicted drug dealers held in Grimsville Prison are placed in cell block A, B, or C according to their city of origin. Danny (who has little knowledge of statistics) was once an inmate in the prison. Now released, he still bears a grudge against the prison authorities. Danny wishes to make up a series of statistics to show that the convicts in the various blocks are treated differently. According to his fictitious sample, the mean number of hours of exercise per week given to the inmates is 10 hours for block A offenders, 20 hours for block B offenders, and 30 hours for block C offenders. Shown below are two fictitious sets of results.

Fictitious study 1:

Block A	Block B	Block C
9	21	30
10	19	29
9	20	31
11	19	29
11	21	31
$\overline{X} = 10$	$\overline{X} = 20$	$\overline{X} = 30$

Fictitious study 2:

Block A	Block B	Block C
18	16	37
16	18	36
10	2	7
2	31	41
4	33	29
$\overline{X} = 10$	$\overline{X} = 20$	$\overline{X} = 30$

a. From simply looking at the numbers, without running any statistical tests, which of the two fictitious studies would you expect to provide stronger backing for Danny's claim? Explain your answer.

b. Calculate the between sum of squares and the within sum of squares for study 1.

c. Calculate the between sum of squares and the within sum of squares for study 2.

d. Calculate the value of eta for each study. How do you account for the difference?

12.7 A researcher takes three independent random samples of young pickpockets and asks them how old they were when they first committed the offense. The researcher wishes to determine whether there are any differences among the three populations from which the samples were drawn—those who have no siblings, those who have one or two siblings, and those with three or more siblings.

Age at first theft:

0 Siblings	1 or 2 Siblings	3+ Siblings
10	14	15
8	15	15
16	15	10
14	13	13
7	12	16
8	9	15

a. Show that the total sum of squares is equal to the between sum of squares plus the within sum of squares.

b. What is the value of eta?

c. Can the researcher reject the null hypothesis on the basis of the differences observed? Run an *F*-test using a 5% significance level. Remember to outline all of the steps of a test of statistical significance, including any violations of assumptions.

12.8 Using independent random sampling, Sophie draws samples from three different populations: psychologists, police officers, and factory workers. She gives each subject a hypothetical case study of a drug dealer who has been found guilty and awaits sentencing. The subjects are then asked to suggest how many years the drug dealer should serve in prison. The results are presented below:

Psychologists	Police	Factory Workers
2	3	5
1	2	6
0	3	4
0	3	8
1	4	7
2.5	1	7
2	1.5	6
1.5	0	2
4	0.5	3
1	7	2

a. Can Sophie conclude that the three populations are different in terms of their attitudes toward punishing convicted drug dealers? Run an *F*-test using a 5% significance level. Remember to outline all of the steps of a test of statistical significance, including any violations of assumptions.

b. Would Sophie's decision be any different if she chose a 1% or a 0.1% level of significance?

c. Calculate the value of eta for the results above. Is the relationship a strong one?

12.9 For the data in Exercise 12.4, run a Kruskal-Wallis test using a 5% level of statistical significance. Remember to outline all of the steps of a test of statistical significance, including any violations of assumptions. Are you able to reject the null hypothesis?

Computer Exercises

In the following discussion for estimating one-way analysis of variance models in SPSS and Stata, you may find it useful to open the data file presented in Table 12.4 in the text (ex12_1.sav or ex12_1.dta) or the corresponding syntax file (Chapter_12.sps or Chapter_12.do).

SPSS

ANOVA

To compute a one-way ANOVA model in SPSS, you will use the ONEWAY command:

```
ONEWAY variable_name(s) BY grouping_variable.
```

The output will present the ANOVA table (discussed in the box on p. 325). If you run this command using the data from Table 12.4 (i.e., ex12_1.sav), it would be

```
ONEWAY age BY crime.
```

The value of the *F*-test reported by SPSS matches that reported on page 321. ANOVA results can also be obtained with the MEANS command:

```
MEANS TABLES=variable_name(s) BY grouping_variable
  /CELLS MEAN COUNT STDDEV
  /STATISTICS ANOVA.
```

where the /CELLS option will generate a table of results that lists the mean, number of cases, and standard deviation for the variable(s) of interest by group category. The /STATISTICS option with ANOVA will generate the ANOVA table produced by ONEWAY (and discussed on p. 325). The ANOVA option will also compute the value of Eta and Eta-squared, which are not available in the ONEWAY command.

Tukey's HSD

The ONEWAY command will perform a wide range of additional calculations on a data file, including Tukey's HSD statistic. To obtain Tukey's HSD statistic, execute the ONEWAY command with the /POSTHOC=TUKEY option:

```
ONEWAY variable_name BY grouping_variable
/POSTHOC=TUKEY.
```

The output presented will contain the ANOVA table that you have already seen and an additional table that lists all possible comparisons of group means.

If you run this command using the data from Table 12.4, the command would be

```
ONEWAY age BY crime
/POSTHOC=TUKEY.
```

Executing this command will reproduce the results from Table 12.10. The three major rows in this table represent the three samples of offenders. Within each major row are two smaller rows that represent contrasts between the groups. So, for example, in the first major row (the embezzlement sample), there are calculations for the mean of this group minus the mean of the second group (the bribery sample) in the first line, followed by calculations for the mean of the first group minus the mean of the third group (the antitrust sample) in the second line. The values for Tukey's HSD reported in the first major row match those reported in Table 12.10. In the second major row (the bribery sample), the second line represents the difference between this group's mean and the mean for the third group (the antitrust sample), and the value for Tukey's HSD again matches that reported in Table 12.10.

Sometimes the labels in the table of results for Tukey's HSD can be confusing, so you will need to pay attention to the lines you are working with. Keep in mind that the variable listed in the first column of each major row has the mean for every other group (listed in the second column) subtracted from its mean.

Kruskal–Wallis Test

The Kruskal–Wallis test is available in SPSS through the use of the NPAR command with the /K-W option:

```
NPAR TESTS
/K-W=variable_name BY grouping_variable (minrange, maxrange).
```

Note the parentheses following the name of the grouping variable. The values to be included here are the values representing the minimum (minrange) and maximum (maxrange) for the grouping variable. Returning to the use of the data in Table 12.4, we would enter the following command:

```
NPAR TESTS
/K-W=age BY crime (1,3).
```

where 1 and 3 represent the bounds of the grouping variable.

The output generated by the NPAR TESTS command will contain two small tables. The first table lists each group or category and its average rank. The second table presents the results for the Kruskal–Wallis test. Note that the value of the test statistic reported by SPSS differs slightly from that reported in the text (SPSS: 13.073; text: 13.038). The reason for this difference was noted in footnote 11: SPSS corrects the calculation of the test statistic by adjusting for ties in rank, and the formula in the text does not make such a correction.

Stata

ANOVA

Similar to many other statistical packages, there are multiple ways of obtaining one-way ANOVA results in Stata. The two most direct commands are **oneway** and **anova**. The **oneway** command provides sufficient information for most purposes. If you are interested in obtaining Tukey's HSD, however, you will need to use the **anova** command. To compute a one-way ANOVA model in Stata, you will use the **oneway** command:

oneway variable_name grouping_variable, **tabulate**

The output will present the ANOVA table. If the **tabulate** option is included on the command line, Stata will generate group means, standard deviations, and counts for the number of cases on the variable of interest. If you omit the **tabulate** option, **oneway** will simply generate an ANOVA table.

To run this command using the data from Table 12.4 (i.e., ex12_1.dta) enter the following:

oneway age crime, **tabulate**

The value of the *F*-test reported by SPSS matches that reported on page 321.

The format for the **anova** command is identical:

anova variable_name grouping_variable

Not surprisingly, the output from this command will be the ANOVA table.

Tukey's HSD

Tukey's HSD statistic is not an option in any of the Stata ANOVA commands. To obtain Tukey's HSD statistic, we must first install a pair of user-written procedures that will use the results from an **anova** command and then compute Tukey's HSD.

To install these user-written procedures, enter the following two commands (one time only for each one):

net install tukeyhsd, from (http://www.ats.ucla.edu/stat/stata/ado/analysis)

net install sg101, from (http://www.stata.com/stb/stb47)

The **tukeyhsd** command will be the procedure we use, but its calculations are based on the other procedure installed (**sg101**).

To obtain Tukey's HSD, we first run a one-way ANOVA using the **anova** command, followed by the **tukeyhsd** command:

anova variable_name grouping_variable

tukeyhsd grouping_variable

For the data in Table 12.4 that we have been working with, the two commands would be

anova age crime

tukeyhsd crime

The output from the running of the **tukeyhsd** command will show only three comparisons, 1 v. 2, 1 v. 3, and 2 v. 3, making the interpretation of the output somewhat simpler than in SPSS, where all possible comparisons are presented. The results are identical to the results from Table 12.10.

Kruskal–Wallis Test

The Kruskal–Wallis test is available in Stata with the **kwallis** command:

kwallis variable_name, **by**(grouping_variable)

Returning to the use of the data in Table 12.4, we would run the following command:

kwallis age, **by**(crime)

The output generated by the **kwallis** command will contain one small table listing the group, number of cases in that group, and rank sum. Below the table, there are chi-square test statistics for both of the methods we have noted: with and without a correction for ties. Consequently, Stata reproduces the value in the text (13.038) that does not correct for ties and the value also estimated by SPSS (13.073) that does correct for ties.

Problems

1. Input the data from Table 12.6 as two variables: bail amount and race (use 1 = non-Hispanic white, 2 = non-Hispanic African American, and 3 = Hispanic of any race).
   a. Reproduce the ANOVA results in the text.
   b. Compute the HSD for these data. What can you conclude about the pairwise comparisons across race categories?
   c. Perform the Kruskal–Wallis test. How do the results from the Kruskal–Wallis test compare to the ANOVA results in part a? Do the results from the Kruskal–Wallis test alter the conclusions obtained through the use of ANOVA?

2. Enter the data from Exercise 12.5. Use one of the ANOVA commands to test for differences in group means.

   a. Write out the assumptions of the test, critical value of the test statistic, value of the computed test statistic, and decision regarding the null hypothesis.

   b. Compute the HSD for each of the group comparisons. What can you conclude about pairwise comparisons for each group?

   c. Use the Kruskal–Wallis test to test for differences in rank order across groups. Write out the assumptions of the test, critical value of the test statistic, value of the computed test statistic, and decision regarding the null hypothesis.

3. Enter the data from Exercise 12.6. Use one of the ANOVA commands to test for differences in group means.

   a. Write out the assumptions of the test, critical value of the test statistic, value of the computed test statistic, and decision regarding the null hypothesis.

   b. Compute the HSD for each of the group comparisons. What can you conclude about pairwise comparisons for each group?

   c. Use the Kruskal–Wallis test to test for differences in rank order across groups. Write out the assumptions of the test, critical value of the test statistic, value of the computed test statistic, and decision regarding the null hypothesis.

4. Open the NYS data file (nys_1.sav, nys_1_student.sav, or nys_1.dta). Carry out the following statistical analyses for each of the research questions in parts a through e:

   a. Does the mean number of thefts valued at $5–$50 vary across academic ability?

   b. Does the mean number of times drunk vary across race?

   c. Does the level of marijuana use vary across amount of contact with delinquent peers?

   d. Does the mean number of attacks on other students vary across victimization experience?

   e. Does the mean number of times cheating on schoolwork vary across grade point average?

      - Use ANOVA to test for differences in group means. For each hypothesis test, write out the assumptions of the test, critical value of the test statistic, value of the computed test statistic, and decision regarding the null hypothesis.

- Compute the HSD for each of the pairwise comparisons. What can you conclude about pairwise comparisons for each research question?
- Use the Kruskal–Wallis test to test for differences in rank order across groups. For each hypothesis test, write out the assumptions of the test, critical value of the test statistic, value of the computed test statistic, and decision regarding the null hypothesis.

5. Open the Pennsylvania Sentencing data file (pcs_ 98.sav or pcs_98.dta). Carry out the following statistical analyses for each of the research questions in parts a through c:

   a. Does the length of incarceration sentence vary across race?

   b. Does the length of incarceration sentence vary across method of conviction?

   c. Does the length of incarceration sentence vary by type of conviction offense?

   - Use ANOVA to test for differences in group means. For each hypothesis test, write out the assumptions of the test, critical value of the test statistic, value of the computed test statistic, and decision regarding the null hypothesis.
   - Compute the HSD for each of the pairwise comparisons. What can you conclude about pairwise comparisons for each research question?
   - Use the Kruskal–Wallis test to test for differences in rank order across groups. For each hypothesis test, write out the assumptions of the test, critical value of the test statistic, value of the computed test statistic, and decision regarding the null hypothesis.