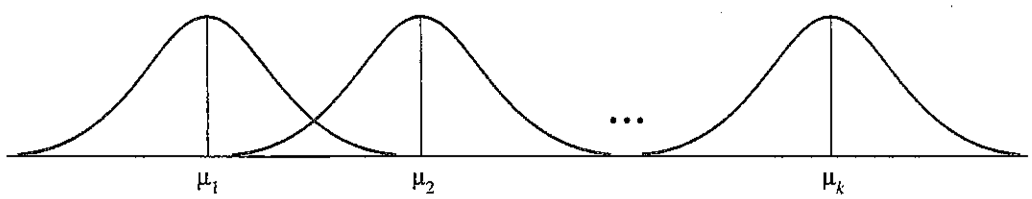
Solution The only way to generate the equivalent of five random samples from the hypothetical populations corresponding to the five insecticides is to use a method called a **randomized assignment.** A fixed number of cotton plants are chosen for treatment, and each is assigned a random number. Suppose that each sample is to have an equal number of measurements. Using a randomization device, you can assign the first n plants chosen to receive insecticide 1, the second n plants to receive insecticide 2, and so on, until all five treatments have been assigned.

Whether by *random selection* or *random assignment,* both of these examples result in a completely randomized design, or one-way classification, for which the analysis of variance is used.

11.5 THE ANALYSIS OF VARIANCE FOR A COMPLETELY RANDOMIZED DESIGN

Suppose you want to compare k population means, $\mu_1, \mu_2, \ldots, \mu_k$, based on independent random samples of size $n_1, n_2, \ldots, n_k$ from normal populations with a common variance σ^2. That is, each of the normal populations has the same shape, but their locations might be different, as shown in Figure 11.2.

FIGURE 11.2
Normal populations with a common variance but different means



Partitioning the Total Variation in an Experiment

Let x_{ij} be the jth measurement ($j = 1, 2, \ldots, n_i$) in the ith sample. The analysis of variance procedure begins by considering the total variation in the experiment, which is measured by a quantity called the **total sum of squares** (**TSS**):

$$\text{Total SS} = \Sigma(x_{ij} - \bar{x})^2 = \Sigma x_{ij}^2 - \frac{(\Sigma x_{ij})^2}{n}$$

This is the familiar numerator in the formula for the sample variance for the entire set of $n = n_1 + n_2 + \cdots + n_k$ measurements. The second part of the calculational formula is sometimes called the **correction for the mean** (**CM**). If we let G represent the *grand total* of all n observations, then

$$\text{CM} = \frac{(\Sigma x_{ij})^2}{n} = \frac{G^2}{n}$$

This Total SS is partitioned into two components. The first component, called the **sum of squares for treatments (SST)**, measures the variation among the k sample means:

$$\text{SST} = \Sigma n_i(\bar{x}_i - \bar{x})^2 = \Sigma\frac{T_i^2}{n_i} - \text{CM}$$

where T_i is the total of the observations for treatment i. The second component, called the **sum of squares for error (SSE)**, is used to measure the pooled variation within the k samples:

$$\text{SSE} = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2$$

This formula is a direct extension of the numerator in the formula for the pooled estimate of σ^2 from Chapter 10. We can show algebraically that, in the analysis of variance,

$$\text{Total SS} = \text{SST} + \text{SSE}$$

Therefore, you need to calculate only two of the three sums of squares—Total SS, SST, and SSE—and the third can be found by subtraction.

Each of the sources of variation, when divided by its appropriate **degrees of freedom,** provides an estimate of the variation in the experiment. Since Total SS involves n squared observations, its degrees of freedom are $df = (n - 1)$. Similarly, the sum of squares for treatments involves k squared observations, and its degrees of freedom are $df = (k - 1)$. Finally, the sum of squares for error, a direct extension of the pooled estimate in Chapter 10, has

$$df = (n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) = n - k$$

Notice that the degrees of freedom for treatments and error are additive—that is,

$$df(\text{total}) = df(\text{treatments}) + df(\text{error})$$

These two sources of variation and their respective degrees of freedom are combined to form the **mean squares** as MS = SS/df. The total variation in the experiment is then displayed in an **analysis of variance** (or **ANOVA**) **table.**

TIP
The column labeled "SS" satisfies: Total SS = SST + SSE.

TIP
The column labeled "df" always adds up to $n - 1$.

ANOVA TABLE FOR k INDEPENDENT RANDOM SAMPLES: COMPLETELY RANDOMIZED DESIGN

Source	df	SS	MS	F
Treatments	$k - 1$	SST	MST = SST/$(k - 1)$	MST/MSE
Error	$n - k$	SSE	MSE = SSE/$(n - k)$	
Total	$n - 1$	Total SS		

where

$$\text{Total SS} = \Sigma x_{ij}^2 - \text{CM}$$
$$= (\text{Sum of squares of all } x\text{-values}) - \text{CM}$$

with

$$\text{CM} = \frac{(\Sigma x_{ij})^2}{n} = \frac{G^2}{n}$$

$$\text{SST} = \Sigma \frac{T_i^2}{n_i} - \text{CM} \qquad \text{MST} = \frac{\text{SST}}{k - 1}$$

$$\text{SSE} = \text{Total SS} - \text{SST} \qquad \text{MSE} = \frac{\text{SSE}}{n - k}$$

and

G = Grand total of all n observations
T_i = Total of all observations in sample i
n_i = Number of observations in sample i
$n = n_1 + n_2 + \cdots + n_k$

EXAMPLE 11.4

In an experiment to determine the effect of nutrition on the attention spans of elementary school students, a group of 15 students were randomly assigned to each of three meal plans: no breakfast, light breakfast, and full breakfast. Their attention spans (in minutes) were recorded during a morning reading period and are shown in Table 11.1. Construct the analysis of variance table for this experiment.

TABLE 11.1 **Attention Spans of Students After Three Meal Plans**

No Breakfast	Light Breakfast	Full Breakfast
8	14	10
7	16	12
9	12	16
13	17	15
10	11	12
$T_1 = 47$	$T_2 = 70$	$T_3 = 65$

Solution To use the calculational formulas, you need the $k = 3$ treatment totals together with $n_1 = n_2 = n_3 = 5$, $n = 15$, and $\Sigma x_{ij} = 182$. Then

$$\text{CM} = \frac{(182)^2}{15} = 2208.2667$$

$$\text{Total SS} = (8^2 + 7^2 + \cdots + 12^2) - \text{CM} = 2338 - 2208.2667 = 129.7333$$

with $(n - 1) = (15 - 1) = 14$ degrees of freedom,

$$\text{SST} = \frac{47^2 + 70^2 + 65^2}{5} - \text{CM} = 2266.8 - 2208.2667 = 58.5333$$

with $(k - 1) = (3 - 1) = 2$ degrees of freedom, and by subtraction,

$$\text{SSE} = \text{Total SS} - \text{SST} = 129.7333 - 58.5333 = 71.2$$

with $(n - k) = (15 - 3) = 12$ degrees of freedom. These three sources of variation, their degrees of freedom, sums of squares, and mean squares are shown in the shaded area of the ANOVA table generated by *MINITAB* and given in Figure 11.3. You will find instructions for generating this output in the "My *MINITAB*" section at the end of this chapter.

FIGURE 11.3
MINITAB output for Example 11.4

One-way ANOVA: Span versus Meal

```
Source  DF      SS     MS      F      P
Meal     2   58.53  29.27   4.93  0.027
Error   12   71.20   5.93
Total   14  129.73

S = 2.436   R-Sq = 45.12%   R-Sq(adj) = 35.97%
                             Individual 95% CIs For Mean
                             Based on Pooled StDev
Level  N    Mean  StDev     --+---------+---------+---------+-------
1      5   9.400  2.302     (---------*--------)
2      5  14.000  2.550                         (---------*--------)
3      5  13.000  2.449                     (---------*--------)
                            --+---------+---------+---------+-------
                            7.5      10.0      12.5      15.0

Pooled StDev = 2.436
```

In Section 11.6, we will present a procedure that you can use to compare all possible pairs of treatment means simultaneously. However, if you have a special interest in a particular mean or pair of means, you can construct confidence intervals using the small-sample procedures of Chapter 10, based on the Student's t distribution. For a single population mean, μ_i, the confidence interval is

$$\bar{x}_i \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n_i}}\right)$$

where $\bar{x}_i$ is the sample mean for the ith treatment. Similarly, for a comparison of two population means—say, μ_i and μ_j—the confidence interval is

$$(\bar{x}_i - \bar{x}_j) \pm t_{\alpha/2}\sqrt{s^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Before you can use these confidence intervals, however, two questions remain:

- How do you calculate s or s^2, the best estimate of the common variance σ^2?
- How many degrees of freedom are used for the critical value of t?

To answer these questions, remember that in an analysis of variance, the mean square for error, MSE, always provides an unbiased estimator of σ^2 and uses information from the entire set of measurements. Hence, it is the best available estimator of σ^2, regardless of what test or estimation procedure you are using. You should *always* use

$$s^2 = \text{MSE} \qquad \text{with } df = (n - k)$$

to estimate σ^2! You can find the positive square root of this estimator, $s = \sqrt{\text{MSE}}$, on the last line of Figure 11.3 labeled "Pooled StDev."

MY TIP

Degrees of freedom for confidence intervals are the *df* for **error.**

COMPLETELY RANDOMIZED DESIGN: (1 − α)100% CONFIDENCE INTERVALS FOR A SINGLE TREATMENT MEAN AND THE DIFFERENCE BETWEEN TWO TREATMENT MEANS

Single treatment mean:

$$\bar{x}_i \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n_i}}\right)$$

Difference between two treatment means:

$$(\bar{x}_i - \bar{x}_j) \pm t_{\alpha/2}\sqrt{s^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

with

$$s = \sqrt{s^2} = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n - k}}$$

where $n = n_1 + n_2 + \cdots + n_k$ and $t_{\alpha/2}$ is based on $(n - k)$ *df*.

EXAMPLE 11.6

The researcher in Example 11.4 believes that students who have no breakfast will have significantly shorter attention spans but that there may be no difference between those who eat a light or a full breakfast. Find a 95% confidence interval for the average attention span for students who eat no breakfast, as well as a 95% confidence interval for the difference in the average attention spans for light versus full breakfast eaters.

Solution For s^2 = MSE = 5.9333 so that $s = \sqrt{5.9333} = 2.436$ with $df = (n - k) = 12$, you can calculate the two confidence intervals:

- For no breakfast:

$$\bar{x}_1 \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n_1}}\right)$$

$$9.4 \pm 2.179\left(\frac{2.436}{\sqrt{5}}\right)$$

$$9.4 \pm 2.37$$

  or between 7.03 and 11.77 minutes.

- For light versus full breakfast:

$$(\bar{x}_2 - \bar{x}_3) \pm t_{\alpha/2}\sqrt{s^2\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}$$

$$(14 - 13) \pm 2.179\sqrt{5.9333\left(\frac{1}{5} + \frac{1}{5}\right)}$$

$$1 \pm 3.36$$

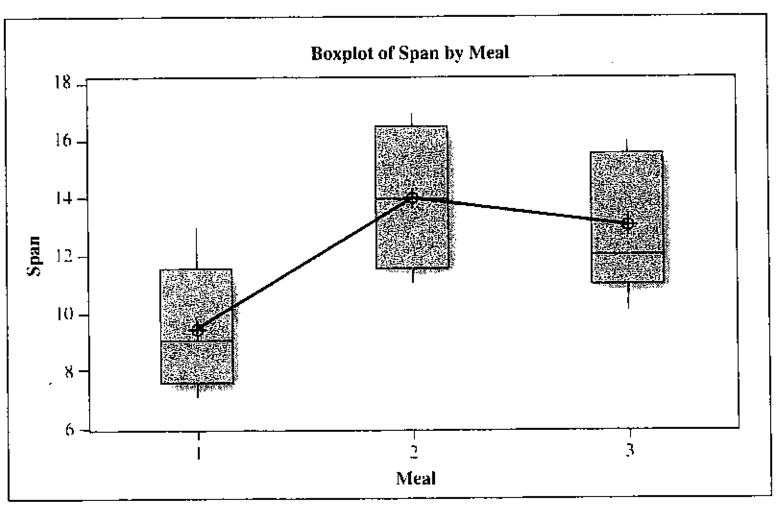
  a difference of between −2.36 and 4.36 minutes.

You can see that the second confidence interval does not indicate a difference in the average attention spans for students who ate light versus full breakfasts, as the researcher suspected. If the researcher, because of prior beliefs, wishes to test the other two possible pairs of means—none versus light breakfast, and none versus full breakfast—the methods given in Section 11.6 should be used for testing all three pairs.

Some computer programs have graphics options that provide a powerful visual description of data and the k treatment means. One such option in the *MINITAB* program is shown in Figure 11.7. The treatment means are indicated by the symbol ⊕ and are connected with straight lines. Notice that the "no breakfast" mean appears to be somewhat different from the other two means, as the researcher suspected, although there is a bit of overlap in the box plots. In the next section, we present a formal procedure for testing the significance of the differences between all pairs of treatment means.

FIGURE 11.7

Box plots for Example 11.6



One option might be to order the sample means from the smallest to the largest and then to conduct t-tests for adjacent means in the ordering. If two means differ by more than

$$t_{\alpha/2}\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

you conclude that the pair of population means differ. The problem with this procedure is that the probability of making a Type I error—that is, concluding that two means differ when, in fact, they are equal—is α for each test. If you compare a large number of pairs of means, the probability of detecting at least one difference in means, when in fact none exists, is quite large.

A simple way to avoid the high risk of declaring differences when they do not exist is to use the **studentized range,** the difference between the smallest and the largest in a set of k sample means, as the yardstick for determining whether there is a difference in a pair of population means. This method, often called **Tukey's method for paired comparisons,** makes the probability of declaring that a difference exists between at least one pair in a set of k treatment means, when no difference exists, equal to α.

Tukey's method for making paired comparisons is based on the usual analysis of variance assumptions. **In addition, it assumes that the sample means are independent and based on samples of equal size.** The yardstick that determines whether a difference exists between a pair of treatment means is the quantity ω (Greek lowercase omega), which is presented next.

YARDSTICK FOR MAKING PAIRED COMPARISONS

$$\omega = q_\alpha(k, df)\left(\frac{s}{\sqrt{n_t}}\right)$$

where

k = Number of treatments
s^2 = MSE = Estimator of the common variance σ^2 and $s = \sqrt{s^2}$
df = Number of degrees of freedom for s^2
n_t = Common sample size—that is, the number of observations in each of the k treatment means
$q_\alpha(k, df)$ = Tabulated value from Tables 11(a) and 11(b) in Appendix I, for $\alpha = .05$ and $.01$, respectively, and for various combinations of k and df

Rule: Two population means are judged to differ if the corresponding sample means differ by ω or more.

Tables 11(a) and 11(b) in Appendix I list the values of $q_\alpha(k, df)$ for $\alpha = .05$ and $.01$, respectively. To illustrate the use of the tables, refer to the portion of Table 11(a) reproduced in Table 11.2. Suppose you want to make pairwise comparisons of $k = 5$ means with $\alpha = .05$ for an analysis of variance, where s^2 possesses 9 df. The tabulated value for $k = 5$, $df = 9$, and $\alpha = .05$, shaded in Table 11.2, is $q_{.05}(5, 9) = 4.76$.

A Partial Reproduction of Table 11(a) in Appendix I; Upper 5% Points

df	2	3	4	5	6	7	8	9	10	11	12
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07	50.59	51.96
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99	14.39	14.75
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61

EXAMPLE

Refer to Example 11.4, in which you compared the average attention spans for students given three different "meal" treatments in the morning: no breakfast, a light breakfast, or a full breakfast. The ANOVA F-test in Example 11.5 indicated a significant difference in the population means. Use Tukey's method for paired comparisons to determine which of the three population means differ from the others.

Solution For this example, there are $k = 3$ treatment means, with $s = \sqrt{\text{MSE}} = 2.436$. Tukey's method can be used, with each of the three samples containing $n_t = 5$ measurements and $(n - k) = 12$ degrees of freedom. Consult Table 11 in Appendix I to find $q_{.05}(k, df) = q_{.05}(3, 12) = 3.77$ and calculate the "yardstick" as

$$\omega = q_{.05}(3, 12)\left(\frac{s}{\sqrt{n_t}}\right) = 3.77\left(\frac{2.436}{\sqrt{5}}\right) = 4.11$$

The three treatment means are arranged in order from the smallest, 9.4, to the largest, 14.0, in Figure 11.8. The next step is to check the difference between every pair of means. The only difference that exceeds $\omega = 4.11$ is the difference between no breakfast and a light breakfast. These two treatments are thus declared significantly different. You cannot declare a difference between the other two pairs of treatments. To indicate this fact visually, Figure 11.8 shows a line under those pairs of means that are not significantly different.

FIGURE 11.8
Ranked means for Example 11.7

None	Full	Light
9.4	13.0	14.0

The results here may seem confusing. However, it usually helps to think of ranking the means and interpreting nonsignificant differences as our inability to distinctly rank those means underlined by the same line. For this example, the light breakfast definitely ranked higher than no breakfast, but the full breakfast could not be ranked higher than no breakfast, or lower than the light breakfast. The probability that we make at least one error among the three comparisons is at most $\alpha = .05$.

MY TIP

If zero is not in the interval, there is evidence of a difference between the two methods.

Most computer programs provide an option to perform **paired comparisons,** including Tukey's method. The *MINITAB* output in Figure 11.9 shows its form of Tukey's test, which differs slightly from the method we have presented. The three intervals that you see in the printout marked "Lower" and "Upper" represent the difference in the two sample means plus or minus the yardstick ω. If the interval contains the value 0, the two means are judged to be not significantly different. You can see that only means 1 and 2 (none versus light) show a significant difference.

FIGURE 11.9
MINITAB output for Example 11.7

```
Tukey's 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of Meal
Individual confidence level = 97.94%

Meal = 1 subtracted from:

Meal  Lower   Center  Upper  -----+---------+---------+---------+----
2      0.493   4.600  8.707           (-----------*-----------)
3     -0.507   3.600  7.707          (-----------*-----------)
                             -----+---------+---------+---------+----
                               -3.5       0.0       3.5       7.0
Meal = 2 subtracted from:

Meal  Lower   Center  Upper  -----+---------+---------+---------+----
3     -5.107  -1.000  3.107  (-----------*-----------)
                             -----+---------+---------+---------+----
                               -3.5       0.0       3.5       7.0
```

As you study two more experimental designs in the next sections of this chapter, remember that, once you have found a factor to be significant, you should use Tukey's method or another method of paired comparisons to find out exactly where the differences lie!

11.6 EXERCISES

BASIC TECHNIQUES

11.19 Suppose you wish to use Tukey's method of paired comparisons to rank a set of population means. In addition to the analysis of variance assumptions, what other property must the treatment means satisfy?

11.20 Consult Tables 11(a) and 11(b) in Appendix I and find the values of $q_\alpha(k, df)$ for these cases:

a. $\alpha = .05, k = 5, df = 7$

b. $\alpha = .05, k = 3, df = 10$

c. $\alpha = .01, k = 4, df = 8$

d. $\alpha = .01, k = 7, df = 5$

11.21 If the sample size for each treatment is n_t and if s^2 is based on 12 *df,* find ω in these cases:

a. $\alpha = .05, k = 4, n_t = 5$

b. $\alpha = .01, k = 6, n_t = 8$

11.22 An independent random sampling design was used to compare the means of six treatments based on samples of four observations per treatment. The pooled estimator of σ^2 is 9.12, and the sample means follow:

$\bar{x}_1 = 101.6$ $\bar{x}_2 = 98.4$ $\bar{x}_3 = 112.3$
$\bar{x}_4 = 92.9$ $\bar{x}_5 = 104.2$ $\bar{x}_6 = 113.8$

a. Give the value of ω that you would use to make pairwise comparisons of the treatment means for $\alpha = .05$.

b. Rank the treatment means using pairwise comparisons.

APPLICATIONS

11.23 Swamp Sites, again Refer to Exercise 11.13 and data set EX1113. Rank the mean leaf growth for the four locations. Use $\alpha = .01$.

11.24 Calcium Refer to Exercise 11.15 and data set EX1115. The paired comparisons option in *MINITAB* generated the output provided here. What do these results tell you about the differences in the population