



Robbie Shone/Science Source

CHAPTER

30

Two-Way Analysis of Variance

Analysis of variance (ANOVA) is a statistical method for comparing the means of several populations based on independent random samples or on the mean responses to several treatments in a randomized comparative experiment. When we compare just two means, we use the two-sample t procedures described in Chapter 21. ANOVA allows comparison of any number of means. The basic form of ANOVA is **one-way ANOVA**, which treats the means being compared as mean responses to different values of a single variable. For example, in Chapter 27 we used one-way ANOVA to compare the mean lengths of three varieties of tropical flowers and the mean number of species per plot after each of three logging conditions.

30.1 Beyond One-Way ANOVA

You should recall or review the big ideas of one-way ANOVA from Chapter 27. One-way ANOVA compares the means $\mu_1, \mu_2, \dots, \mu_I$ of I populations based on samples of sizes n_1, n_2, \dots, n_I from these populations.

- The **conditions for ANOVA** require *independent random samples* from each of the I populations (or a randomized comparative experiment with I treatments), *Normal distributions* for the response variable in each population, and a *common standard deviation* σ in all populations. Fortunately, ANOVA inference is quite robust against moderate violations of the Normality and common standard deviation conditions.

In this chapter, we cover...

- 30.1 Beyond one-way ANOVA
- 30.2 Two-way ANOVA: Conditions, main effects, and interaction
- 30.3 Inference for two-way ANOVA
- 30.4 Some details of two-way ANOVA*

- One-way ANOVA gives a single test for the null hypothesis that all the population means are the same against the alternative hypothesis that not all are the same.
- ANOVA works by comparing how far apart the sample means are relative to the variation among individual observations in the same sample. The test statistic is the ANOVA *F* statistic

ANOVA *F* statistic

$$F = \frac{\text{variation among the sample means}}{\text{variation among individuals in the same sample}}$$

F distribution

Tukey's pairwise multiple comparisons

- The *P*-value comes from an *F* distribution.
- The *F* test can be combined with a formal *follow-up analysis* using **Tukey's pairwise multiple comparisons** to determine which of the pairs of means differ and to estimate how large these differences are.

Here is an example that illustrates one-way ANOVA.

EXAMPLE 30.1 Logging in the Rain Forest



Fletcher & Baylis/Science Source

STATE: How does logging in a tropical rain forest affect the forest in later years? Researchers compared forest plots in Borneo that had never been logged (Group A) with similar plots nearby that had been logged one year earlier (Group B) and eight years earlier (Group C). Although the study was not an experiment, the authors explain why we can consider the plots to be randomly selected. Table 30.1 displays data on the number of tree species found in each plot.¹ Is there evidence that the mean numbers of species in the three groups differ?

PLAN: Describe how the three samples appear to differ, check the conditions for ANOVA, and carry out the one-way ANOVA *F* test to compare the three population means.

SOLVE: The stemplots in Figure 30.1 show that the distributions are irregular in shape, as is common for small samples. None of the distributions are strongly skewed,

TABLE 30.1 Counts of tree species in forest plots in Borneo

Group A Never Logged	Group B Logged One Year Ago	Group C Logged Eight Years Ago
22	11	17
18	11	4
22	14	18
20	7	14
15	18	18
21	15	15
13	15	15
13	12	10
19	13	12
13	2	
19	15	
15	8	

Group A (never)	Group B (1 year)	Group C (8 years)
2	2 0	2
3	3	3
4	4	4 0
5	5	5
6	6	6
7	7 0	7
8	8 0	8
9	9	9
10	10	10 0
11	11 00	11
12	12 0	12 0
13 000	13 0	13
14	14 0	14 0
15 00	15 000	15 00
16	16	16
17	17	17 0
18 0	18 0	18 00
19 00	19	19
20 0	20	20
21 0	21	21
22 00	22	22

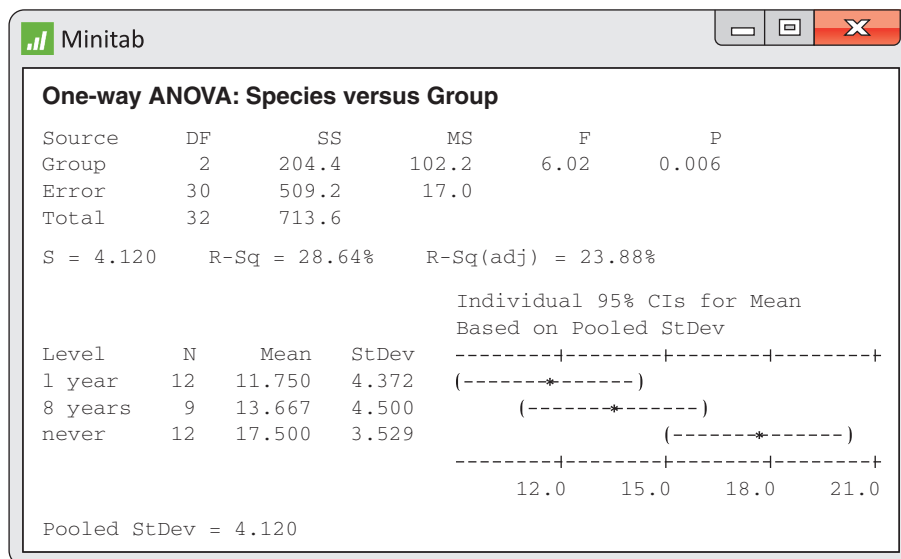
FIGURE 30.1

Side-by-side stemplots comparing the counts of tree species in three groups of forest plots, from Table 30.1.

and there are no extreme outliers. The unlogged Group A plots tend to have more species than the two logged groups and also show less variation among plots. The Minitab ANOVA output in Figure 30.2 shows that the standard deviations satisfy our rule of thumb (page 647) that the largest (4.5) is no more than twice the smallest (3.529). The output also shows that the mean number of species in Group A is higher than in Groups B and C. The ANOVA F test is highly significant, with P -value $P = 0.006$.

CONCLUDE: The data provide strong evidence ($F = 6.02$, $P = 0.006$) that the mean number of species per plot is not the same in all three groups. It appears that the mean species count is higher in unlogged plots.

Minitab

**FIGURE 30.2**

Minitab ANOVA output, for the logging study of Example 30.1.

Example 30.1, and one-way ANOVA in general, compares mean responses for several values of just one explanatory variable. In Example 30.1, that variable is “how long ago this plot was logged.” Suppose that we have data on *two* explanatory variables: say, how long ago the plot was logged and whether it is located in a river bottom or in the highlands. There are now six groups formed by the combinations of time since logging and location. Figure 30.3 illustrates the combinations of these two factors. One-way ANOVA will still tell us if there is evidence that mean species counts in these six types of plot differ. But we want more: Does location matter? Does time since logging matter? And do these two variables *interact*? That is, does the effect of logging change when we move from river bottoms to highlands? Perhaps trees regrow faster in river bottoms so that logging has less effect there than in the highlands. To answer these questions, we must extend ANOVA to take into account the fact that the six groups are formed from two explanatory variables. This is done with a **two-way ANOVA**.

two-way ANOVA

FIGURE 30.3
The treatments in the two-factor logging study. Combinations of values of the two factors form six treatments.

		Time Since Logging		
		Never logged	One year	Eight years
Plot Location	River bottom	Group 1	Group 2	Group 3
	Highlands	Group 4	Group 5	Group 6

In the next section, we examine the conditions for a two-way ANOVA and the interpretation of main effects and interaction. This is followed by formal inference for two-way ANOVA in Section 30.3.

APPLY YOUR KNOWLEDGE

In each of these exercises, identify the two explanatory variables, the number of treatments, and the response. Then use a diagram like Figure 30.3 to display the explanatory variables and treatments

- 30.1 Violence and Sex in Television Programs.** Do violence and sex in television programs help sell products in advertisements? Subjects were randomly assigned to watch one of four types of TV shows: (1) neither sex nor violence in the content code, (2) violence but no sex in the content code, (3) sex but no violence in the content code, and (4) both sex and violence in the content code. For each TV show, the original advertisements were replaced with the same set of 12 advertisements. Subjects were not told the purpose of the study, but were instead told that the researchers were studying attitudes toward TV shows. After viewing the show, subjects received a surprise memory test to check their recall of the products advertised.²
- 30.2 Policy Justification: Pragmatic vs. Moral.** How does a leader’s justification of his or her organization’s policy affect support for the policy? This study compared a moral, pragmatic, and ambiguous justification for both

public and private policies. As an example of a public policy, subjects read a politician’s proposal to fund a retirement planning agency. The moral justification was the importance of retirees “to live with dignity and comfort,” the pragmatic was “to not drain public funds,” and the ambiguous was “to have sufficient funds.” For an example of a private policy, subjects read about a CEO’s plan to provide healthy meals for the employees. The moral justification was increased access to meals “should improve our employees’ well-being,” the pragmatic was “to improve our employees’ productivity,” and the ambiguous was “to improve the status-quo.” Each subject was randomly assigned to one of the justifications for either a public or private policy. After reading the proposal, subjects answered several questions measuring support for the policy proposal to create a support score, higher values indicating greater support.³

30.2 Two-Way ANOVA: Conditions, Main Effects, and Interaction

One-way analysis of variance compares the mean responses from any set of populations or experimental treatments when the responses satisfy the ANOVA conditions. Often, however, a sample or experiment has some design structure that leads to more specific questions than those answered by the one-way ANOVA *F* test or by Tukey pairwise comparisons. As described in the previous section, we may wish to compare treatments that are combinations of values of two explanatory variables, or two **factors** in the language of experimental design. Here is an example that we will use throughout this section.

factor

EXAMPLE 30.2 Effects of TV Advertising

What are the effects of repeated exposure to an advertising message? The answer may depend both on the length of the ad and on how often it is repeated. To investigate this question, assign undergraduate students (the *subjects*) to view a 40-minute television program that includes ads for a digital camera. Some subjects see a 30-second commercial; others, a 90-second version. The same commercial appears either one, three, or five times during the program.



This experiment has two *factors*: length of the commercial, with two values; and repetitions, with three values. The six combinations of one value of each factor form six *treatments*. The subjects are divided at random into six groups, one for each treatment:

		Variable C Repetitions		
		1 Time	3 Times	5 Times
Variable R	30 seconds	Group 1	Group 2	Group 3
Length	90 seconds	Group 4	Group 5	Group 6

This is a *two-way layout*, with values of one factor forming rows and values of the other forming columns. After watching the TV program, all the subjects fill out a questionnaire that produces an “intention to buy” score with values between zero and 100. This is the *response variable*.

two-way ANOVA conditions

To analyze data from such a study, we impose some additional conditions. Here are the **two-way ANOVA conditions** that will govern our work on this topic:

crossed designs

- 1. We have *responses for all combinations* of values of the two factors (all six cells in Example 30.2). No combinations are missing in our data. In general, call the two explanatory variables *R* and *C* (for Row and Column). Variable *R* has *r* different values and variable *C* has *c* different values. The study compares all *rc* combinations of these values. These are called **crossed designs**.
- 2. In an observational study, we have *independent SRSs* from each of the *rc* populations. If the study is an experiment, the available subjects are allocated at random among all *rc* treatments. That is, we have a *completely randomized design*. We also allow some *randomized block designs* in which we have an SRS from each of *r* populations and the subjects in each SRS are then separately allocated at random among the same *c* treatments.⁴ We met block designs in Chapter 9 (see page 244).
- 3. The response variable has a *Normal distribution* in each population. The population mean responses may differ, but all *rc* populations have a *common standard deviation* σ .
- 4. We have the *same number of individuals* *n* in each of the *rc* treatment groups or samples. These are called **balanced designs**.

balanced designs

As always, the design of the study is the most important condition for statistical inference. Conditions 1 and 2 describe the designs covered by two-way ANOVA. These designs are very common. Condition 3 lists the usual ANOVA conditions on the distribution of the response variable. ANOVA inference is reasonably robust against violations of these conditions.

When you design a study, you should try to satisfy the fourth condition; that is, you should choose equal numbers of individuals for each treatment. The *F* tests are most robust against violation of the “common standard deviation” condition when the subject counts are equal or close to equal in either a one-way or a two-way layout. In addition, for the two-way layout, there is an even stronger reason to prefer balanced designs. If the numbers of individuals differ among treatments (an unbalanced design), several alternative analyses of the data are possible. These analyses answer different sets of questions, and you must decide which questions you want to answer. All the sets of questions and all the analyses collapse to just one in the balanced case. This makes interpreting your data much simpler.

To understand the questions that two-way ANOVA answers, return to the advertising study in Example 30.2. In this section, we will assume that we know the actual population mean responses for each treatment. That is, we deal with an ideal situation in which we don’t have to worry about random variation in the mean responses.

EXAMPLE 30.3 Main Effects with No Interaction

Here again is the two-way layout of the advertising study in Example 30.2. The numbers in the six cells are now made-up values of the population mean responses to the six treatments:

		Variable C Repetitions		
		1 Time	3 Times	5 Times
Variable R	30 seconds	30	45	50
Length	90 seconds	40	55	60

The mean “intention to buy” scores increase with more repetitions of the camera ad and also increase with the length of the ad. The means increase *by the same amount* (10 points) when we move from 30 seconds to 90 seconds, *no matter how many times the ad is shown*. Turning to the other variable, the effect of moving from one to three repetitions is the same (15 points) for both 30-second and 90-second ads, and the effect of moving from three to five repetitions is also the same (5 points). Because the result of changing the value of one variable is the same for all values of the other variable, we say that there is *no interaction* between the two variables.

Now average the mean responses for 30 seconds and for 90 seconds. The average for 30 seconds is

$$\frac{30 + 45 + 50}{3} = \frac{125}{3} = 41.7$$

and the average for 90-second ads is

$$\frac{40 + 55 + 60}{3} = \frac{155}{3} = 51.7$$

Because the averages for the two lengths differ, we say that there is a *main effect* for length. Similarly, average the mean responses for each number of repetitions. They are 35 for one repetition, 50 for three, and 55 for five. So changing the number of times the ad is shown has an “on the average” effect on the response. There is a *main effect* for repetitions.

Figure 30.4 plots the cell means from Example 30.3. The two solid lines joining one, three, and five repetitions for 30-second and for 90-second ads are *parallel*. This reflects the fact that the mean response always increases by 10 points when we move from 30-second to 90-second ads, no matter how many times the ad is shown. *Parallel lines in a plot of means show that there is no interaction*. It doesn’t matter which variable you choose to place on the horizontal axis.

To see the main effect of repetitions, look at the average response for 30 and 90 seconds at each number of repetitions. This average is the dotted line in the plot. It changes as we move from one to three to five repetitions. *A variable has a main effect when the average response differs for different values of the variable*. “Average” here means averaged over all the values of the other variable. A main effect of repetitions



Rocked by 'roids

How do baseball fans react to accusations of steroid use by some of the game’s most high-profile players? Rafael Palmeiro, one of four players with 3000 hits and 500 home runs, was suspended for 10 days in 2005 for a steroid violation after insisting at a congressional hearing that he never used steroids. When Barry Bonds hit his 756th home run to break Hank Aaron’s record, many fans thought his accomplishment was tainted by allegations of steroid use. Should an asterisk be placed behind the names of these players in record books? Public reaction depends on geographic region, gender, age, race, and many other factors. Any careful study of what fans think will include several factors.

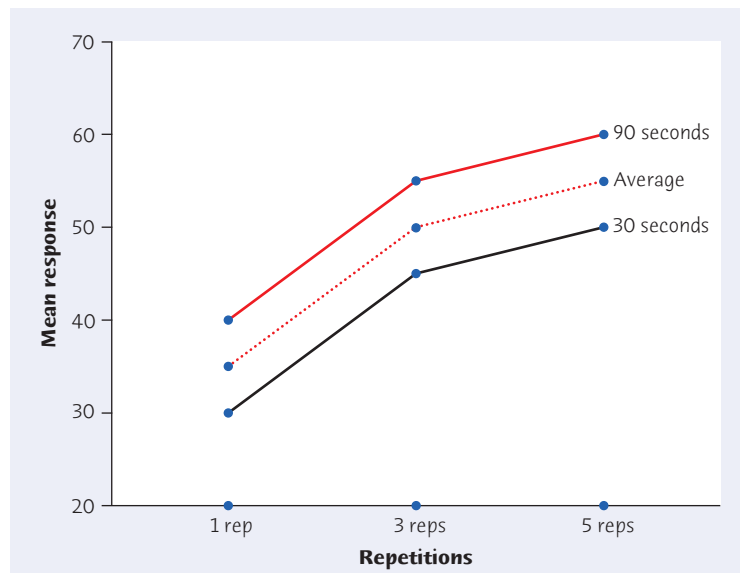


FIGURE 30.4

Plot of the means, from Example 30.3. In addition to the means themselves, the plot displays their average for each number of repetitions as a dotted line. The parallel lines show that there is no interaction between the two factors.

is present in Figure 30.4 because the dotted “average” line is not horizontal. A main effect for length is also present, but it can’t be seen directly in the plot.

EXAMPLE 30.4 Interactions and Main Effects

Now look at this different set of mean responses for the advertising study:

		Variable C Repetitions		
		1 Time	3 Times	5 Times
Variable R	30 seconds	30	45	50
Length	90 seconds	40	45	40

Figure 30.5 plots these means, with the means for 30-second ads connected by solid black lines and those for 90-second ads by solid red lines.

These means reflect the case in which subjects grow annoyed when the ad is both longer and repeated more often. The mean responses for the 30-second ad are the same as in Example 30.3. But showing the 90-second ad three times increases the mean response by only 5 points, and showing it five times drops the mean back to 40. There *is* an interaction between length and repetitions: the difference between 30 and 90 seconds changes with the number of repetitions, so that the solid black and red lines in Figure 30.5 are not parallel.

There is still a main effect of repetitions because the average response (dotted line in Figure 30.5) changes as we move from one to three to five repetitions. What about the effect of length? The average overall values of repetitions for 30-second ads is

$$\frac{30 + 45 + 50}{3} = \frac{125}{3} = 41.7$$

For 90-second ads, this average is the same,

$$\frac{40 + 45 + 40}{3} = \frac{125}{3} = 41.7$$

On the average over all numbers of repetitions, changing the length of the ad has no effect. There is *no main effect* for length.

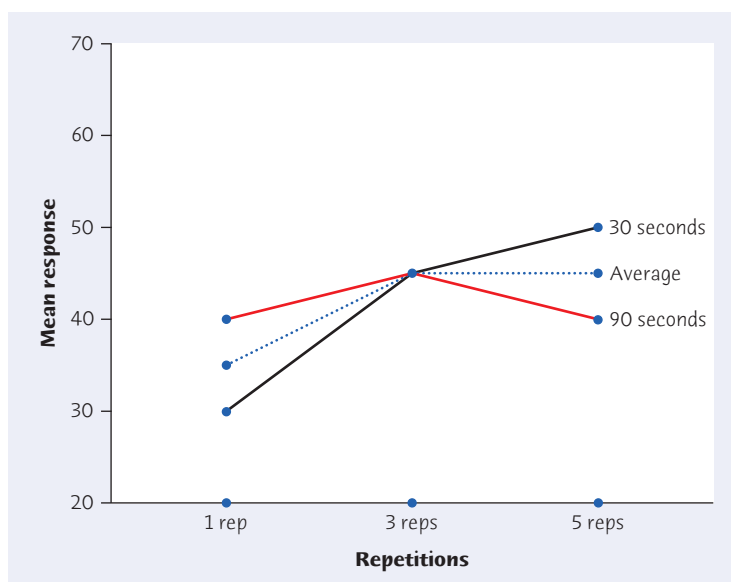


FIGURE 30.5

Plot of the means, from Example 30.4, along with the average for each number of repetitions. The lines are not parallel, so there is an interaction between length and number of repetitions.

Interactions and Main Effects

An **interaction** is present between factors R and C in a two-way layout if the change in mean response when we move between two values of R is different for different values of C . (We can interchange the roles of R and C in this statement.)

A **main effect** for factor R is present if, when we average the responses for a fixed value of R over all values of C , we do not get the same result for all values of R .



Main effects may have little meaning when interaction is present. After all, interaction says that the effect of changing one of the variables is different for different values of the other variable. The main effect, as an “on the average” effect, may not tell us much. In Example 30.4, there is no main effect for the length of the camera ad. But length certainly matters—Figure 30.5 shows that there is little point in paying for 90-second ads if the same ad will be shown several times during the program. That’s the interaction of length with number of repetitions.

EXAMPLE 30.5 Which Effect Is More Important?

There are no simple rules for interpreting results from two-way ANOVA when strong interaction is present. You must look at plots of means and think. Figure 30.6 displays two different mean plots for a study of the effects of classroom conditions on the performance of normal and hyperactive schoolchildren. The two conditions are “quiet” and “noisy,” where the noisy condition is actually the usual environment in elementary school classrooms.

There is an interaction: normal children perform a bit better under noisy conditions, but hyperactive children perform slightly less well under noisy conditions. The interaction is exactly the same size in the two plots of Figure 30.6. To see this, look at the slopes of the “Normal” lines in the two plots: they are the same. The slopes of the two “Hyperactive” lines are also the same. So the size of the gap between normal and hyperactive changes by the same amount when we move from quiet to noisy in both plots, even though the gap is much larger in Figure 30.6(b).

In Figure 30.6(a), this interaction is the most important conclusion of the study. Both main effects are small: normal children do a bit better than hyperactive children, for example, but not a great deal better on the average.



Aniel Skelley/Getty Images

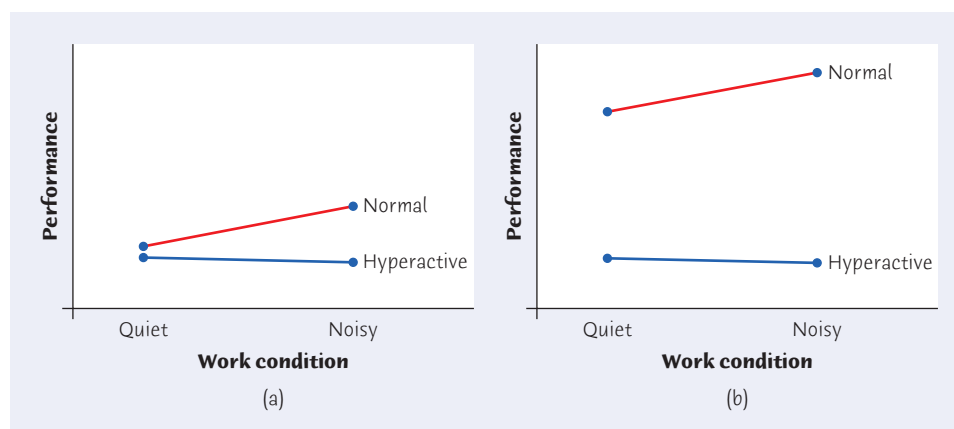


FIGURE 30.6

Plots of two sets of means for a study comparing the performance of normal and hyperactive children under two conditions, for Example 30.5.

In Figure 30.6(b), the main effect of “hyperactive or not” is the big story. Normal children perform much better than hyperactive children in both environments. The interaction is still there, but it is not very important in the face of the large difference in average performance between hyperactive and normal children.

In this section, we pretended that we knew the population means so that we could discuss patterns without needing statistical inference. In practice, we don’t know the population means. However, plotting the sample means for all groups is an essential part of data analysis for a two-way layout. Look for interaction and main effects just as we did in this section. Of course, in real data you will almost never find exactly parallel lines representing exactly no interaction. Two-way ANOVA inference helps guide you because it assesses whether the interaction in the data is statistically significant. We are now ready to introduce two-way ANOVA inference.

Macmillan Learning Online Resources

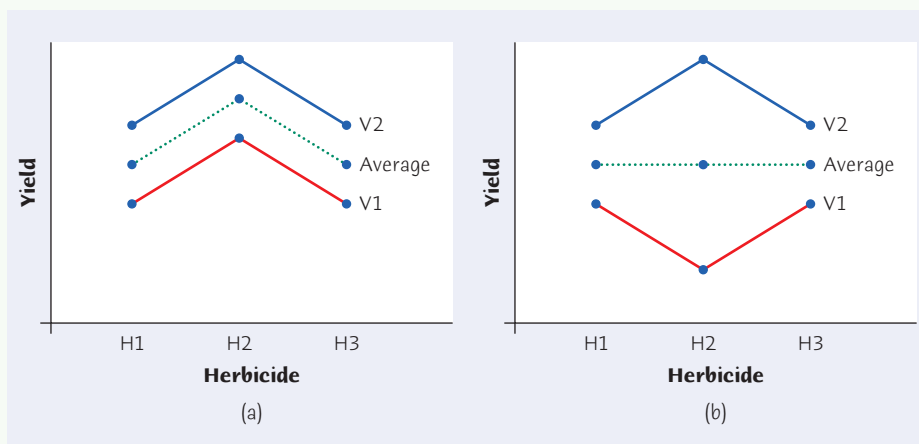
- The StatBoards video, *Two-Way ANOVA: Main Effects and Interaction*, interprets plots of the means in a two-way ANOVA. Examples with and without interaction are provided and the meaning of the main effects is examined in both cases.

APPLY YOUR KNOWLEDGE

Figure 30.7 shows two different plots of means for a two-way study that compares the yields of two varieties of soybeans (V1 and V2) when three different herbicides (H1, H2, and H3) are applied to the fields. Exercises 30.3 and 30.4 ask you to interpret these plots.

FIGURE 30.7

Plots of two sets of possible means for a study in which the two factors are soybean variety and type of herbicide, for Exercises 30.3 and 30.4. The plots also show the average for each herbicide (dotted line).



30.3 Recognizing Effects. Consider the mean responses plotted in Figure 30.7(a).

- Is there an interaction between soybean variety and herbicide type? Why or why not?
- Is there a main effect of herbicide type? Why or why not?
- Is there a main effect of soybean variety? Why or why not?

30.4 Recognizing Effects. Consider the mean responses plotted in Figure 30.7(b).

- Is there an interaction between soybean variety and herbicide type? Why or why not?

- (b) Is there a main effect of herbicide type? Why or why not?
 (c) Is there a main effect of soybean variety? Why or why not?

30.5 Assigning Subjects. Exercise 30.2 describes a two-factor experiment in which there are six treatments. How are the subjects to be assigned to the treatments? This exercise describes two methods for assigning subjects to the treatments. Only one of the two methods will satisfy our conditions for the analysis of a two-way ANOVA described in this chapter.

- (a) Suppose there are 300 subjects and they are assigned at random to the six treatments, 50 subjects to each treatment. There are four conditions given for the two-way ANOVA on page 30-6. Are Conditions 1, 2, and 4 satisfied? Explain.
 (b) Here is how the study was actually conducted. There were 150 subjects and experimenters assigned 50 at random to each of the three policy justifications. Each subject then read both the public and private proposals for the policy justification to which they were assigned. As in part (a), there are again 50 subjects reading the policy proposals for each of the six treatments. Are Conditions 1, 2, and 4 satisfied? Explain.

30.6 Angry Women, Sad Men. What are the relationships among the portrayal of anger or sadness, gender, and the degree of status conferred? Sixty-eight subjects were randomly assigned to view a videotaped interview in which either a male or a female professional described feeling either anger or sadness. The people being interviewed (we'll call them the "targets") wore professional attire and were ostensibly being interviewed for a job. The targets described an incident in which they and a colleague lost an account and, when asked by the interviewer how it made them feel, responded either that the incident made them feel angry or that it made them feel sad. The subjects were divided into four groups; each group evaluated one of the four types of interviews.⁵ After watching the interview, subjects evaluated the target on a composite measure of status conferral that included items assessing how much status, power, and independence the target deserved in his or her future job. The measure of status ranged from 1 = none to 11 = a great deal. Here are the summary statistics:

Treatment	<i>n</i>	\bar{x}	<i>s</i>
Males expressing anger	17	6.47	2.25
Females expressing anger	17	3.75	1.77
Males expressing sadness	17	4.05	1.61
Females expressing sadness	17	5.02	1.80

- (a) Display the four treatment means in a two-way layout similar to those given in Exercises 30.3 and 30.4.
 (b) Plot the means and discuss the interaction and the two main effects.

30.3 Inference for Two-Way ANOVA

Inference for two-way ANOVA is, in many ways, similar to inference for one-way ANOVA. Here is a brief outline:

1. Find and plot the group sample means. Study the plot to understand the interaction and main effects. Do data analysis to check the conditions for ANOVA.

2. Use software for basic ANOVA inference. There are now three F tests with three P -values, which answer the questions:

Is the interaction statistically significant?

Is the main effect for variable R statistically significant?

Is the main effect for variable C statistically significant?

We will illustrate two-way ANOVA inference with several examples. In the first example, the interaction is both small and insignificant, so that the message is in the main effects.

EXAMPLE 30.6 Computer-Assisted Instruction



COMPINST

STATE: A study in education concerned computer-aided instruction in the use of “Blissymbols” for communication. Blissymbols are pictographs (think of Egyptian hieroglyphs) sometimes used to help learning-disabled children. Typically abled schoolchildren were randomly assigned to treatment groups. There are four groups in a two-way layout:

		Learning Style	
		Active	Passive
Placement	Before	Group 1	Group 2
	During	Group 4	Group 3

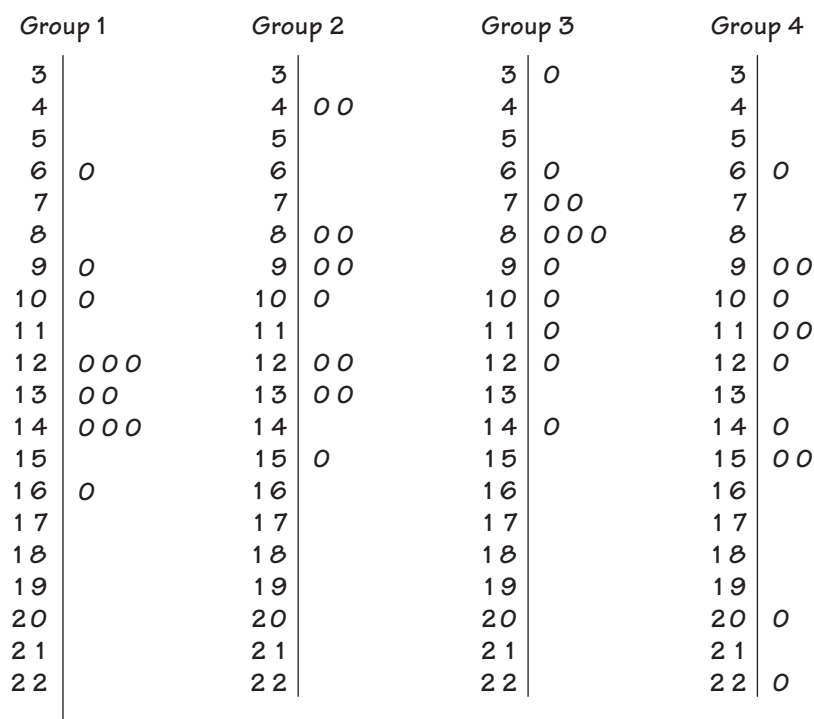
The two factors are the learning style (active or passive) of the lesson and the placement of the material to be learned (Blissymbols presented before compounds, or Blissymbols and compounds shown during a joint presentation). The response variable is the number of symbols recognized correctly, out of 24, in a test taken after the lesson. Table 30.2 displays the data.⁶

PLAN: Plot the sample means and discuss interaction and main effects. Check the conditions for ANOVA inference. Use two-way ANOVA F tests to determine the significance of interaction and main effects.

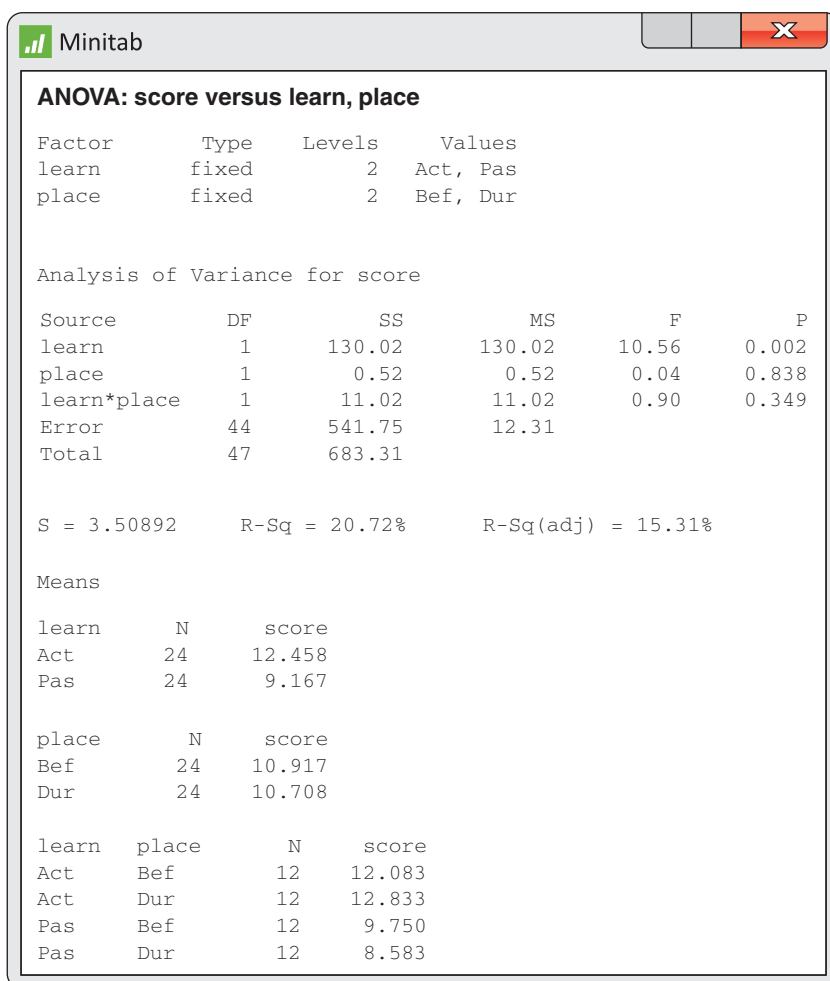
SOLVE: This is a balanced completely randomized experimental design. Figure 30.8 displays side-by-side stemplots of the scores for the 12 subjects in each group. Figure 30.9 shows two-way ANOVA output from Minitab. The stemplots

TABLE 30.2 Test scores in an education study

Group 1	Group 2	Group 3	Group 4
14	4	11	12
14	4	8	22
16	9	8	9
14	8	7	14
6	15	14	20
10	8	9	15
12	9	7	9
13	13	8	10
12	13	12	11
12	12	10	11
13	12	6	15
9	10	3	6

**FIGURE 30.8**

Side-by-side stemplots comparing the counts of correct answers for subjects in the four treatment groups, for Example 30.6.

Minitab**FIGURE 30.9**

Two-way ANOVA output from Minitab for the computer-assisted instruction study, for Example 30.6.

Minitab

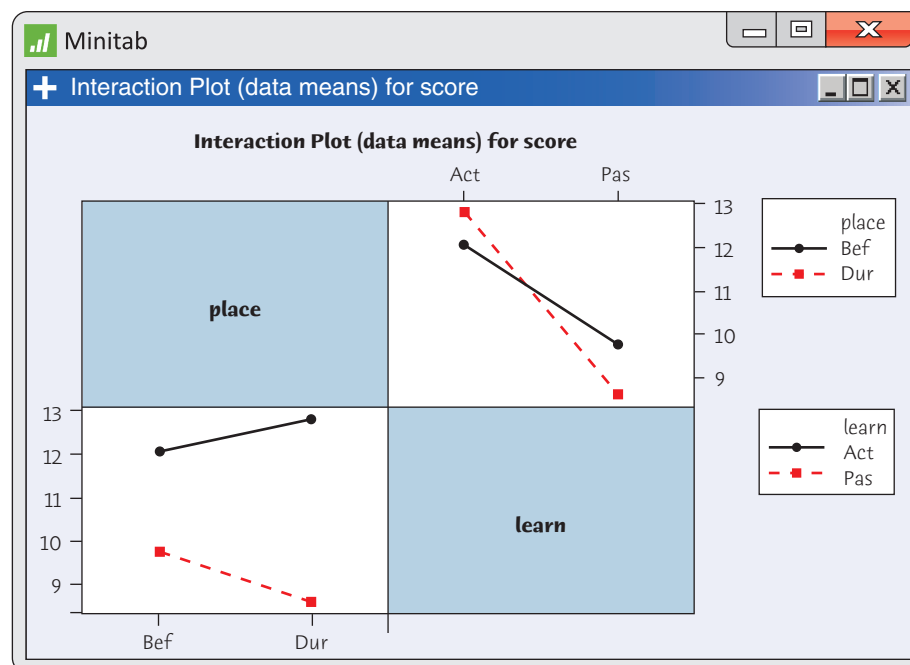


FIGURE 30.10

Minitab plots of the group means from the computer-assisted learning study, for Example 30.6. The two plots use the same four means. They differ only in the choice of which variable to mark on the horizontal axis.

show no departures from Normality strong enough to rule out ANOVA. (One observation, the 6 in Group 1, is slightly outside the range allowed by the $1.5 \times IQR$ rule. This would not be unusual among 48 Normally distributed observations.) You can check that the group standard deviations satisfy our rule of thumb that the largest (4.648) is no more than twice the smallest (2.678). We can proceed with ANOVA inference.

Figure 30.10 shows the plots of means produced by Minitab. The two plots display the same four sample means; Minitab helpfully gives a plot with each variable marked on the horizontal axis. The plots are easy to interpret: the interaction and the main effect of placement are both small, and the main effect of learning style is quite large. The three F tests in the Minitab output substantiate what the plots of means show: interaction ($P = 0.349$) and placement ($P = 0.838$) are not significant, but learning style ($P = 0.002$) is highly significant.

CONCLUDE: Educators know that active learning almost always beats passive learning. This is the only significant effect that appears in these data. In particular, placement of material has very little effect under either style of learning.

The second example illustrates the situation in which there is significant interaction, but main effects are larger and more important. Think of Figure 30.6(b).

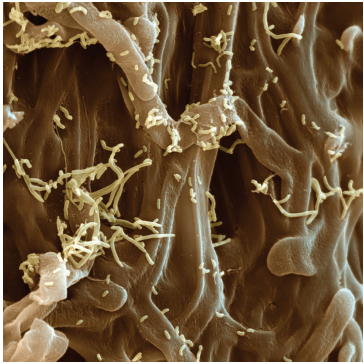
EXAMPLE 30.7 Mycorrhizal Colonies and Plant Nutrition



STATE: Mycorrhizal fungi are present in the roots of many plants. This is a symbiotic relationship, in which the plant supplies nutrition to the fungus and the fungus helps the plant absorb nutrients from the soil. An experiment compared the effects of

adding nitrogen fertilizer to two genetic types of tomato plants: a normal variety that is susceptible to mycorrhizal colonies and a mutant that is not. Nitrogen was added at rates of 0, 28, or 160 kilograms per hectare (kg/ha). Here is the two-way layout for the six treatment combinations:

		Tomato Type	
		Mutant	Normal
Nitrogen	0 kg/ha	Group 1	Group 4
	28 kg/ha	Group 2	Group 5
	160 kg/ha	Group 3	Group 6



Six plants of each type were assigned at random to each amount of fertilizer. The response variables describe the levels of nutrients in a plant after 19 weeks, when the tomatoes are fully ripe. We will look at one response, the percent of phosphorus in the plant. Table 30.3 contains the data.⁷



PLAN: Plot the sample means and discuss interaction and main effects. Check the conditions for ANOVA inference. Use two-way ANOVA *F* tests to determine the significance of interaction and main effects.

SOLVE: This is a randomized block design. We are willing to assume that each of our two sets of tomato plants is an SRS from its genetic type. We consider the blocks (genetic types) as one of the factors in analyzing the data because we are interested in comparing the two genetic types. In addition to the two-way ANOVA, we might consider separate one-way ANOVAs for mutant and normal types to draw separate conclusions about the effect of fertilizer on each type.

TABLE 30.3 Percent of phosphorus in tomato plants

Mutant			Normal		
Group	Fertilizer	Percent Phosphorus	Group	Fertilizer	Percent Phosphorus
1	0	0.29	4	0	0.64
1	0	0.25	4	0	0.54
1	0	0.27	4	0	0.53
1	0	0.24	4	0	0.52
1	0	0.24	4	0	0.41
1	0	0.20	4	0	0.43
2	28	0.21	5	28	0.41
2	28	0.24	5	28	0.37
2	28	0.21	5	28	0.50
2	28	0.22	5	28	0.43
2	28	0.19	5	28	0.39
2	28	0.17	5	28	0.44
3	160	0.18	6	160	0.34
3	160	0.20	6	160	0.31
3	160	0.19	6	160	0.36
3	160	0.19	6	160	0.37
3	160	0.16	6	160	0.26
3	160	0.17	6	160	0.27

FIGURE 30.11
Minitab plots of the group means for the study of phosphorus levels in tomatoes, for Example 30.7.

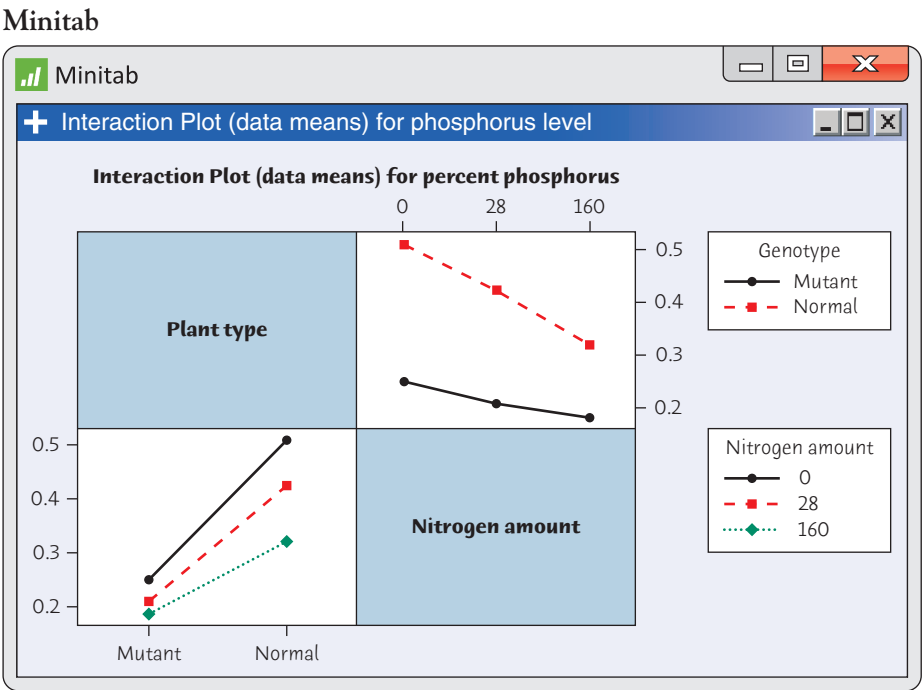


Figure 30.11 displays Minitab's plots of the six sample means. The lines are not parallel, so interaction is present. The interaction is rather small compared with the main effects. The main effect of type is expected: normal plants have higher phosphorus levels than the mutants at all levels of fertilization because they benefit from symbiosis with the fungus. The main effect of fertilizer is a bit surprising: phosphorus level goes down as the level of nitrogen fertilizer increases.

Examination of the data (we don't show the details) finds no outliers or strong skewness. But the largest sample standard deviation (0.08329 in Group 4) is much larger than twice the smallest (0.01472 in Group 3). As the number of treatment groups increases, even samples from populations with exactly the same standard deviation are more likely to produce sample standard deviations that violate our "twice as large" rule of thumb. (Think of comparing the shortest and tallest person among more and more people.) So our rule of thumb is often conservative for two-way ANOVA. Nonetheless, ANOVA inference may not give correct P -values for these data. The P -values for the three two-way ANOVA F tests are $P = 0.008$ for interaction and $P < 0.001$ for both main effects. These agree with the mean plots and are so small that even if not accurate, they strongly suggest significance.

CONCLUDE: Normal plants, with their mycorrhizal colonies, have higher phosphorus levels than mutants that lack such colonies. Nitrogen fertilizer actually reduces phosphorus levels in both types of plants. The reduction is stronger for normal plants, but this interaction is not very large in practical terms.

Finally, here is an example in which the main interest in the study is the interaction between the two factors.

EXAMPLE 30.8 The Effect of Product Density and Ambient Scent on Consumer Anxiety



STATE: Although retail stores overflowing with merchandise can make consumers anxious, minimally stocked spaces can have the same effect. Researchers investigated whether the use of ambient scents can reduce anxiety by creating feelings

of openness in a crowded environment or coziness in a minimally stocked environment. Participants were invited to a lab that simulated a retail environment that was either jam-packed (high density) or nearly empty (low density). For each of these two product densities, the lab was infused with one of three scents: (1) a scent associated with spaciousness such as the seashore, (2) a scent associated with an enclosed space like the smell of firewood, and (3) no scent at all. Two hundred ninety-eight participants were randomly assigned to the six treatments, 33 to each treatment. Participants were told the purpose of the study was to evaluate retail merchandise and the environment. Consumers evaluated several products and their level of anxiety was measured using the State-Trait Anxiety Inventory.⁸

PLAN: Plot the sample means and discuss interaction and main effects. Check the conditions for ANOVA inference. Use two-way ANOVA F tests to determine the significance of interaction and main effects.

SOLVE: This is a balanced completely randomized experiment with six treatments. Figure 30.12 shows Minitab's plots of the sample means. The plots are all in the expected direction. For a low-density site, an intimate scent reduces anxiety relative to no scent, while a spacious scent increases anxiety. For the high-density site the opposite is true. A spacious scent reduces anxiety, while an intimate scent increases anxiety. In the plot, this appears as an interaction with the high-density plot showing increasing anxiety and the low-density plot showing decreasing anxiety.

Although we don't show the details, ANOVA is justified. There are no outliers or strong skewness. The largest sample standard deviation (0.69 for high density and intimate scent) is less than twice the smallest (0.37 for low density and intimate scent).

Figure 30.13 contains Minitab's ANOVA output, with the means and standard deviations for the six groups. Although not highly significant, the interaction is the main story. Because the variability in the anxiety scores with treatments is large relative to the size of the differences among the six treatment means, the P -value for the test for interaction is not that small despite the strong pattern in the plot in Figure 30.12 and the large sample size. However, the fact that all the means follow the hypothesized pattern makes the results convincing despite the P -value of 0.068.



A Momentary Lapse

Beware of distracted drivers. A recent study showed that putting on makeup, chatting on a cell phone, eating breakfast, reading, and reaching for a moving object significantly increase the risk of being involved in an accident. About 80 percent of crashes occur within three seconds of a driver's distraction. As you might guess, the risk depends on the type of distraction. As usual, other factors also play a role: for example, young drivers are four times more likely to have crashes and near-crashes as drivers over the age of 35.



DENSITY

Minitab

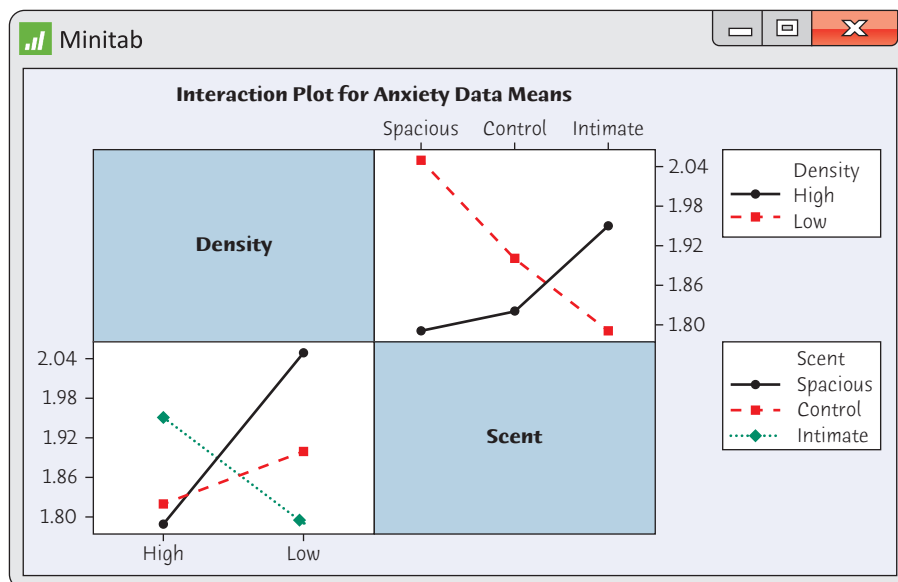
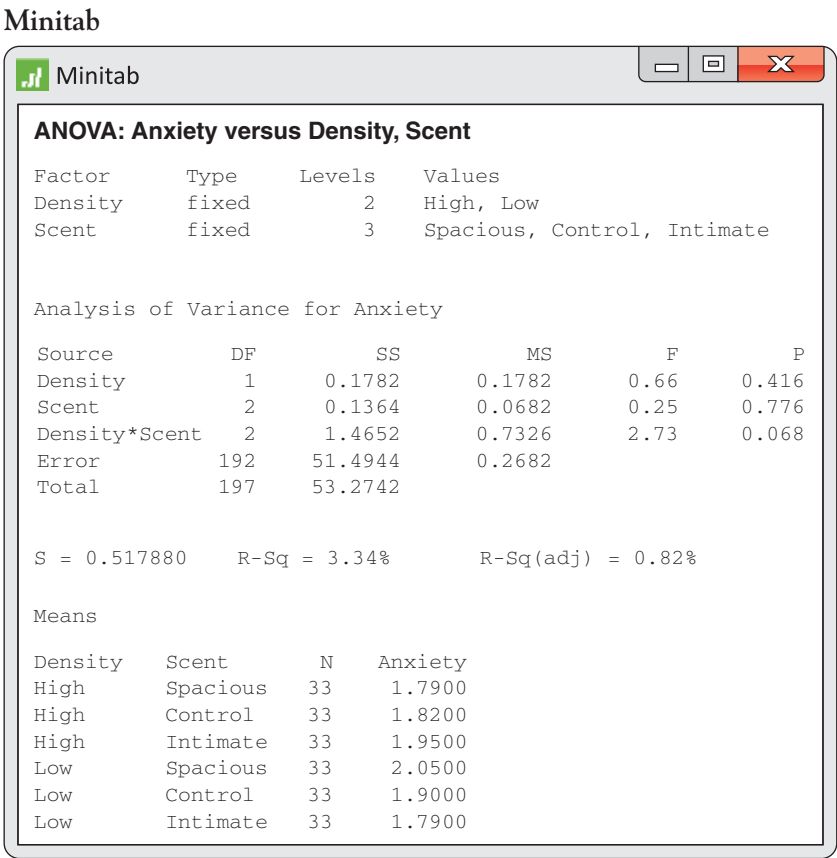


FIGURE 30.12

Minitab plots of the mean levels of anxiety for the combinations of product density and ambient scent, for Example 30.8.

FIGURE 30.13
Two-way ANOVA output from Minitab for the study of the effect of product density and ambient scent on anxiety, for Example 30.8.




CONCLUDE: The data support the hypothesis that the use of ambient scents can reduce anxiety by creating feelings of openness in a crowded environment or coziness in a minimally stocked environment.

Macmillan Learning Online Resources

- The StatBoards videos, *Two-Way ANOVA without Interaction* and *Two-Way ANOVA with Interaction*, provide the details of inference in two-way ANOVA, including the two-way ANOVA *F* tests and the interpretation of the results.


APPLY YOUR KNOWLEDGE

- 30.7 Hooded Rats: Social Play Times.** How does social isolation during a critical developmental period affect the behavior of hooded rats? Psychology students assigned 24 young female rats at random to either isolated or group housing, then similarly assigned 24 young male rats. This is a randomized block design with the gender of the 48 rats as the blocking variable and housing type as the treatment. Later, the students observed the rats at play in a group setting and recorded data on three types of behavior (object play, locomotor play, and social play).⁹ The data file records the time (in seconds) that each rat devoted to social play during the observation period.  SCLTIME
- (a) Make a plot of the four group means. Is there a large interaction between gender and housing type? Which main effect appears to be more important?
 - (b) Verify that the conditions for ANOVA inference are satisfied.




ARCO/H. Reinhard/AGE Fotostock

- (c) Give the complete two-way ANOVA table. What are the F statistics and P -values for interaction and the two main effects? Explain why the test results confirm the tentative conclusions you drew from the plot of means.

30.8 Hooded Rats: Social Play Counts. The researchers who conducted the study in Exercise 30.7 also recorded the number of times each of three types of behavior (object play, locomotor play, and social play) occurred. The data file contains the counts of social play episodes by each rat during the observation period. Use two-way ANOVA to analyze the effects of gender and housing.  SCLCOUNT



30.9 Metabolic Rates in Caterpillars. Professor Harihuko Itagaki and his students have been measuring metabolic rates in tobacco hornworm caterpillars (*Manduca sexta*) for years. The researchers do not want the metabolic rates to depend on which analyzer they use to obtain the measurements. They therefore make six repeated measurements on each of three caterpillars with each of three analyzers.¹⁰  CATRPLRS

- (a) Use software to give the two-way ANOVA table. The researchers used caterpillar as a blocking variable because metabolic rates vary from individual to individual. These three caterpillars are not of interest in themselves. They represent the larger population of caterpillars of this species. We should therefore avoid inference about the main effect of caterpillars.
- (b) Explain why the researchers will be unhappy if there is a significant interaction between analyzer and caterpillar. What does your analysis show about the interaction?
- (c) Is there a significant effect for analyzer? Do you think the researchers will be happy with this result?

30.4 Some Details of Two-Way ANOVA*

All ANOVA F statistics work on the same principle: compare the variation due to the effect being tested with a benchmark level of variation that would be present even if that effect were absent. The three F tests for two-way ANOVA use the same benchmark as the one-way ANOVA F test—namely, the variation among individual responses within the same treatment group.

In two-way ANOVA, we have two factors (explanatory variables) that form treatments in a two-way layout. Factor R has r values and Factor C has c values, so there are rc treatments. The same number n of subjects are assigned to each treatment. The two-way layout that results is as follows:

		Column Factor C			
		1	2	...	c
Factor R	1	n subjects	n subjects	...	n subjects
	2	n subjects	n subjects	...	n subjects
	\vdots	\vdots	\vdots		\vdots
	r	n subjects	n subjects	...	n subjects

The number of treatments is $I = rc$
 The total number of observations is $N = rcn$

*This optional material requires the optional section on some details of ANOVA on pages 658–660.

FIGURE 30.14
Compare the sums of squares in these one-way and two-way ANOVA outputs, for the study in Example 30.6.

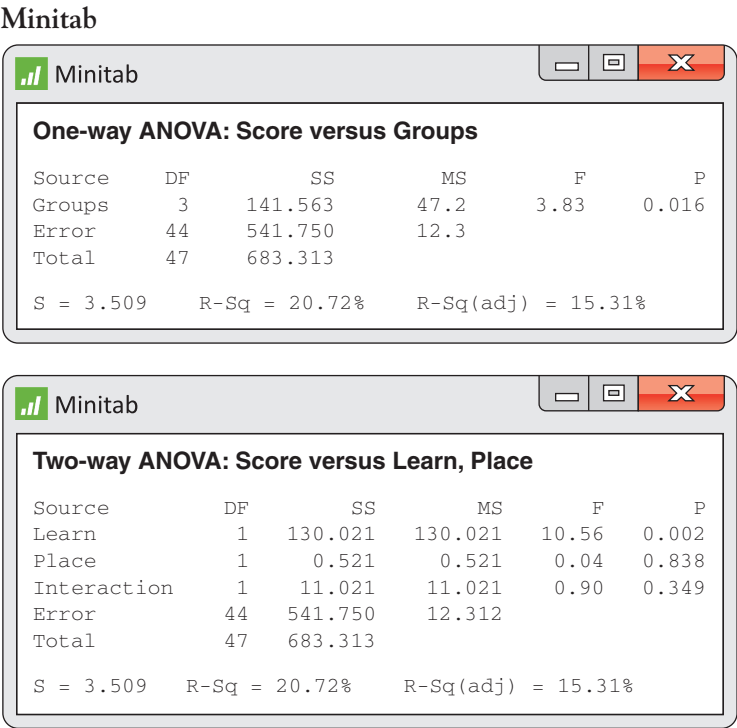


Figure 30.14 presents both one-way and two-way ANOVA output for the same data, from the computer-assisted education study in Example 30.6. The one-way analysis just compares the means of rc treatments, ignoring the two-way layout. In discussing one-way ANOVA, we called the number of treatments I . Now $I = rc$. The two-way analysis takes into account that each treatment is formed by combining a value of R with a value of C .

In both settings, analysis of variance breaks down the overall variation in the observations into several pieces. The overall variation is expressed numerically by the **total sum of squares**

$$SSTO = \sum (\text{individual observation} - \text{mean of all observations})^2$$

where the sum runs over all N individual observations. If we divide $SSTO$ by $N - 1$, we get the variance of the observations. So $SSTO$ is closely related to a familiar measure of variability. Because $SSTO$ uses just the N individual observations, it is the same for both one-way and two-way analyses. You can see in Figure 30.14 that $SSTO = 683.313$ for these data.

One-way ANOVA. We saw in Chapter 27 that the one-way ANOVA F test compares the variation among the I treatment means with the variation among responses to the same treatment. If the means vary more than we would expect based on the variation among subjects who receive the same treatment, that's evidence of a difference among the mean responses in the I populations.

Let's give a bit more detail. One-way ANOVA breaks down the total variation into the sum of two parts:

$$\text{total variation among responses} = \text{variation among treatment means} + \text{variation among responses to the same treatment}$$

$$\text{total sum of squares} = \text{sum of squares for groups} + \text{sum of squares for error}$$

$$SSTO = SSG + SSE$$

Formulas for the sums of squares SSG and SSE appear on page 658 as the numerators of the mean squares MSG and MSE, but we won't concern ourselves with the algebra. Remember that "error" is the traditional term in ANOVA for variation among observations. It doesn't imply that a mistake has been made. In the one-way output in Figure 30.14, you see that the breakdown for these data is

$$\begin{aligned} \text{SSTO} &= \text{SSG} + \text{SSE} \\ 683.313 &= 141.563 + 541.750 \end{aligned}$$

The one-way ANOVA F test is formed in two stages (page 659):

1. Divide each sum of squares by its *degrees of freedom* to get the *mean squares* MSG for groups and MSE for error,

$$\text{MSG} = \frac{\text{SSG}}{I - 1} \quad \text{MSE} = \frac{\text{SSE}}{N - I}$$

2. The one-way ANOVA F statistic compares MSG to MSE,

$$F = \frac{\text{MSG}}{\text{MSE}}$$

Find the P -value from the F distribution with $I - 1$ and $N - I$ degrees of freedom.

The **ANOVA table** in the output reports sums of squares, their degrees of freedom, mean squares, and the F statistic with its P -value.

ANOVA table

Two-way ANOVA. Now look at the two-way analysis of variance table in Figure 30.14:

- The total sum of squares and the error sum of squares are the same as in the one-way analysis.
- The sum of squares for groups in a one-way analysis is the sum of the three sums of squares for main effects and interaction in a two-way analysis.

This is the heart of two-way analysis of variance: break down the variation among the rc groups into three parts: variation due to the main effect of Factor R , variation due to the main effect of Factor C , and variation due to interaction between the two factors. Each type of variation is measured by a sum of squares. The formulas for the two main effects sums of squares are similar to that for the one-way sum of squares for groups, but we will again ignore the algebraic details. The interaction sum of squares is best thought of as what's left over: the variation among treatments that isn't explained by the two main effects. In symbols,

$$\begin{aligned} \text{total sum of squares} &= \text{sum of squares for main effect of Factor } R \\ &\quad + \text{sum of squares for main effect of Factor } C \\ &\quad + \text{sum of squares for interaction between } R \text{ and } C \\ &\quad + \text{sum of squares for error} \end{aligned}$$

$$\text{SSTO} = \text{SSR} + \text{SSC} + \text{SSRC} + \text{SSE}$$

Each of these sums of squares has a degrees of freedom, and these also break down in the same way:

$$\begin{aligned} \text{total df} &= \text{df for main effect of Factor } R \\ &\quad + \text{df for main effect of Factor } C \\ &\quad + \text{df for interaction between } R \text{ and } C \\ &\quad + \text{df for error} \\ rcn - 1 &= (r - 1) + (c - 1) + (r - 1)(c - 1) + rc(n - 1) \end{aligned}$$

You can check that the total degrees of freedom in the line above are $N - 1$ and the degrees of freedom for error are $N - I$, the same as for one-way ANOVA. The one-way degrees of freedom for groups are the sum of the degrees of freedom for the three two-way effects.

EXAMPLE 30.9 Comparing One-Way and Two-Way ANOVA



COMPINST

The data behind Figure 30.14 appear in Table 30.2. The two factors are R = placement of Blissymbol elements and C = learning style. Factor R has $r = 2$ values: Before and During. Factor C has $c = 2$ values: Active and Passive. There are $I = 4$ treatments and $n = 12$ subjects assigned to each treatment, resulting in $N = 48$ observations.

The total degrees of freedom are $N - 1 = 47$. The degrees of freedom for error are $N - I = 48 - 4 = 44$. In the one-way analysis, the degrees of freedom for groups are $I - 1 = 3$. The two-way analysis breaks this into degrees of freedom $r - 1 = 1$ for Factor R , $c - 1 = 1$ for Factor C , and $(r - 1)(c - 1) = 1 \times 1 = 1$ for interaction.

Here are the two breakdowns of the total variation and the degrees of freedom that appear in Figure 30.14:

One-Way			Two-Way		
Sums of Squares		df	Sums of Squares		df
SSG	141.563	3	SSR	130.021	1
			SSC	0.521	1
			SSRC	11.021	1
SSE	541.750	44	SSE	541.750	44
SSTO	683.313	47	SSE	683.313	47

The neat breakdown of SSG into three effects depends on the balance of the two-way layout. It doesn't hold when the counts of observations are not the same for all treatments. That's why two-way ANOVA for unbalanced data is more complicated and harder to interpret than for balanced data.

Two-way ANOVA F tests. Finally, form three F statistics exactly as in the one-way setting.

1. Divide each sum of squares by its *degrees of freedom* to get the *mean squares* for the three effects and for error:

$$MSR = \frac{SSR}{r - 1} \quad MSC = \frac{SSC}{c - 1} \quad MSRC = \frac{SSRC}{(r - 1)(c - 1)} \quad MSE = \frac{SSE}{N - I}$$

2. The three F statistics compare the mean squares for the three effects with MSE.

Two-Way ANOVA F Tests

The F statistics for the three types of treatment effects in two-way ANOVA are:


For the main effect of Factor R , $F = \frac{MSR}{MSE}$ with dfs $r - 1$ and $N - I$

For the main effect of Factor C, $F = \frac{MSC}{MSE}$ with dfs $c - 1$ and $N - I$


For the interaction of R and C, $F = \frac{MSRC}{MSE}$ with dfs $(r - 1)(c - 1)$ and $N - I$

In all cases, large values of F are evidence against the null hypothesis that the effect is not present in the populations.

APPLY YOUR KNOWLEDGE

30.10 The Effect of Product Density and Ambient Scent on Consumer Anxiety, continued. In Example 30.8, a two-way ANOVA examined the effects of product density and ambient scent on consumer anxiety in a simulated retail environment. Start your work in this exercise with the two-way ANOVA table in Figure 30.13 (page 30-18).  **DENSITY**

- Explain how the sums of squares from the two-way ANOVA table can be combined to obtain the one-way ANOVA sum of squares for the six groups (SSG). What is the value of SSG?
- Give the degrees of freedom, mean square (MSG), and F statistic for testing for the effect of groups.
- Is there a significant effect of group on the level of consumer anxiety as measured by the State-Trait Anxiety Inventory? Give and interpret the P -value in the context of this experiment. Is the P -value surprising given the results found in Example 30.8? Explain briefly.
- Use software to carry out one-way ANOVA of consumer anxiety on groups. Verify that your results in parts (a), (b), and (c) agree with the software output.

30.11 Hooded Rats: Social Play Counts, continued. In Exercise 30.8, you carried out two-way ANOVA for a study of the effect of social isolation on hooded rats. The response variable is the count of social play episodes during an observation period. Start your work in this exercise with your two-way ANOVA table from Exercise 30.8.  **SCLCOUNT**

- How many treatment groups are there in this experiment?
- Starting from the two-way ANOVA table, create a one-way ANOVA table to examine the effect of the four groups.
- Is there a significant group effect? Give the appropriate hypotheses, test statistic, P -value, and conclusion in the context of this experiment.

CHAPTER 30 SUMMARY

Chapter Specifics

- Two-way analysis of variance (ANOVA)** compares the means of several populations formed by combinations of two factors R and C in a **two-way layout**.
- The **conditions for ANOVA** state that we have an **independent SRS** from each population (or a completely randomized or randomized block experimental design), that each population has a **Normal distribution**, and that all populations have the **same standard deviation**. In this chapter, we consider only examples that satisfy the additional conditions that the design producing the data is **crossed** (all combinations of the factors are present) and **balanced** (all factor combinations are represented by the same number of individuals).

- A factor has a **main effect** if the mean responses for each value of that factor, averaged over all values of the other factor, are not the same. The two factors **interact** if the effect of moving between two values of one factor is different for different values of the other factor. **Plot the treatment mean responses** to examine main effects and interaction.
- There are three ANOVA **F tests**: for the null hypotheses of no main effect for Factor R, no main effect for Factor C, and no interaction between the two factors.

Statistics in Summary

Here are the most important skills you should have acquired from reading this chapter.

A. Recognition

1. Recognize the two-way layout, in which we have a quantitative response to treatments formed by combinations of values of two factors.
2. Recognize when comparing mean responses to the treatments in a two-way layout is helpful in understanding data.
3. Recognize when you can use two-way ANOVA to compare means. Check the data production, the presence of outliers, and the sample standard deviations for the groups you want to compare. Look for data production designs that are crossed and balanced.

B. Interpreting Two-Way ANOVA

1. Plot the sample means for the treatments. Based on your plot, describe the main effects and interaction that appear to be present.
2. Decide which effects are most important in practice. Pay particular attention to whether a large interaction makes one or both main effects less meaningful.
3. Use software to carry out two-way ANOVA inference. From the *P*-values of the three *F* tests, learn which effects are statistically significant.

Link It

The multiple regression models in Chapter 29 extend the simple linear regression model, allowing us to study the effect of several predictors on the mean of a quantitative response. Similarly, the two-way ANOVA extends the one-way ANOVA described in Chapter 27, allowing us to study the effect of two factors on the mean of a quantitative response. In the two-way ANOVA, we have specific questions about the two factors: what are the main effects of each factor, and do the two factors interact? These questions can be answered by examining appropriate plots of the means and the *F* tests in the ANOVA table. In the presence of strong interaction, the main effects may be of little interest.

Macmillan Learning Online Resources

If you are having difficulty with any of the sections of this chapter, this online resource should help prepare you to solve the exercises at the end of this chapter:

- LearningCurve provides you with a series of questions about the chapter that adjust to your level of understanding.

CHECK YOUR SKILLS

30.12 The purpose of two-way ANOVA is to learn about

- (a) the means of two populations.
- (b) the variances of two populations.
- (c) the combined effects of two factors on a quantitative response.

30.13 A two-way ANOVA compares two exercise programs (A and B) and two exercise frequencies (twice a week and four times a week). The response variable is weight lost after eight weeks of exercise. An interaction is present in the data when

- (a) the mean weight loss is not the same for Program A and Program B.
- (b) the mean weight loss under Program A is different for twice-weekly and four-times-weekly subjects.
- (c) the difference between the mean weight losses for Programs A and B is not the same for twice-weekly and four-times-weekly subjects.

30.14 When the F test for interaction in a two-way ANOVA is significant,

- (a) the main effects are never meaningful and should be ignored.
- (b) the main effects must be interpreted with caution.
- (c) interpret each main effect just as you would in a one-way ANOVA.

A student project measured the increase in the heart rates of fellow students when they stepped up and down for three minutes to the beat of a metronome. The explanatory variables are step height (Lo = 5.75 inches, Hi = 11.5 inches) and metronome beat (Slow = 14 steps/minute, Med = 21 steps/minute, Fast = 28 steps/minute). The subject's heart rate was measured for 20 seconds before and after stepping. The response variable is the increase in heart rate during exercise.¹¹ Use the ANOVA table in Figure 30.15 to answer the Questions 30.15 through 30.19.



HEARTRATE

30.15 How many treatment groups were there for this experiment?

- (a) 3 (b) 5 (c) 6

30.16 What is the value of the test statistic for interaction between frequency and height?

- (a) 218.4 (b) 109.2 (c) 0.97 (d) 0.393

30.17 What is the estimate for the common standard deviation σ ?

- (a) 218.4 (b) 109.2 (c) 10.61

30.18 How many students were in the study?

- (a) 29 (b) 30 (c) It cannot be determined from the ANOVA table.

30.19 Which statement provides the best summary of the interaction between frequency and height?

- (a) The interaction is significant at the 1% level but not at the 5% level.
- (b) The interaction is significant at the 5% level but not at the 1% level.
- (c) The interaction is not significant at either the 1% or the 5% level.

Minitab

Source	DF	SS	MS	F	P
Freq	2	4048.8	2024.4	17.99	0.000
Height	1	1920.0	1920.0	17.07	0.000
Interaction	2	218.4	109.2	0.97	0.393
Error	24	2700.0	112.5		
Total	29	8887.2			

S = 10.61 R-Sq = 69.62% R-Sq(adj) = 63.29%

FIGURE 30.15

Two-way ANOVA table, for Exercises 30.15 through 30.19.

CHAPTER 30 EXERCISES

30.20 Hooded rats: Object play times. Exercise 30.7 (page 30-18) describes an experiment to study the effects of social isolation on the behavior of hooded rats. You have analyzed the effects on social play. Now look at another response variable, the time that a rat spends in object play during an observation period. The data file records the time (in seconds) that each rat devoted to object play.



OBJTIME


- (a) Make a plot of the four group means. Is there a large interaction between gender and housing type? Which main effect appears to be more important?


(b) Verify that the conditions for ANOVA inference are satisfied.

(c) Give the complete two-way ANOVA table. What are the F statistics and P -values for interaction and the two main effects? Explain why the test results confirm the tentative conclusions you drew from the plot of means.


30.21 Hooded rats: Object play counts. The researchers who conducted the study in the previous exercise also recorded the number of times each of three types of




behavior (object play, locomotor play, and social play) occurred. The data file contains the counts of object play episodes for each rat during the observation period. Carry out a complete analysis of the effects of gender and housing type.  OBJCOUNT

30.22 Herbicide and corn hybrids. Genetic engineering has produced new corn hybrids that resist the effects of herbicides. This allows more effective control of weeds, because herbicides don't damage the corn. A study compared the effects of the herbicide glufosinate on a number of corn hybrids. The percents of necrosis (leaf burn) 10 days after application of glufosinate for several application rates (kilograms per hectare) and three corn hybrids, two resistant and one not, are provided in the data file.¹²  CORNHYB

- Construct a plot of means to examine the effects of application rate and hybrid and their interaction.
- Are the conditions for ANOVA inference satisfied? Explain.


30.23 Girls' cross-country times. Ten runners each completed 10 five-kilometer races for the Thomas Worthington High School girls' cross-country team during the 2004 season. The data file contains each race time in minutes.¹³  GIRLSRUN

- Prepare side-by-side boxplots to compare the distributions of race times for the 10 runners. Identify the best runner on the team.
- Use runner as a blocking variable to see if there are significant differences in the overall mean race times for the different meets. Identify the appropriate parameters, state the null and alternative hypotheses, calculate the test statistic and *P*-value, and state your conclusion in context.

30.24 Assembling furniture: Times. Are there sex differences in furniture assembly speed, and does it depend on the type of instructions? Men and women assembled a kitchen trolley from IKEA using either step-by-step instructions or a diagram of the finished product. Eighty volunteer students, 40 women and 40 men, from the University of Tromsø participated in the study. The 40 men were randomly divided into two groups of 20 each, with one group given the step-by-step instructions and the other group given the diagram of the finished product. The same was done for the women. Subjects had 30 minutes to complete the task. For those subjects who completed the task in under 30 minutes, the actual assembly times were used and those subjects that did not finish were assigned a time of 30 minutes.¹⁴  FURNTIME

- Make a plot of the four group means. Is there a large interaction between gender and instruction type? What does the plot suggest about the two main effects?
- Verify that the conditions for ANOVA inference are satisfied.

- Give the complete two-way ANOVA table. What are the *F* statistics and *P*-values for interaction and the two main effects? Explain why the test results confirm the tentative conclusions you drew from the plot of means.

30.25 Assembling furniture: Scores. The researchers who conducted the study in the previous exercise also recorded an assembly score for each subject on the scale 0, 1, . . . , 10. Two raters independently assigned an assembly score to each subject, and differences between the scores were resolved between the judges to give a single score. A score of 10 represents a correctly assembled trolley. Minor errors such as the wired shelf fastened upside down or having a few unused screws received a score of 9. Middle scores had an error such as an unattached rail or top or a combination of two errors with one being minor. The lower scores were reserved for more serious errors.  FURNSCORE

- The assembly scores took values from 0 to 10. Nevertheless, ANOVA is appropriate in this situation. Briefly explain why.
- Carry out a complete analysis of the effects of gender and instruction type.

30.26 Fourth-graders composing music. The Orff xylophone is often used in teaching music to children because it has removable bars that allow the teacher to present different options to the students. An education researcher used the Orff xylophone to examine the effect of tonality (pentatonic or harmonic minor) and number of xylophone bars (five or 10) on the ability of fourth-graders to compose melodies that they could play repeatedly.¹⁵



ArcLight/Alamy

- Twelve children were randomly assigned to each combination of tonality and bar count. Give the two-way layout for this experiment.
- Judges listened to tapes of the children's work and assigned scores for several aspects of the melodies.

One response variable measured the extent to which children generated new musical ideas in consecutive five-second intervals of their melodies. Here is the two-way ANOVA table for this variable (the publication does not give the group means):

Source	df	Sum of Squares	<i>F</i>	<i>P</i>
Tonality	1	44.08	0.43	0.52
Bar count	1	705.33	6.81	0.01
Interaction	1	50.02	0.48	0.49
Error	44	4556.54		

Comment on the significance of main effects and interaction. Then make a recommendation for teachers who want to use the Orff xylophone to encourage children to generate melodies with new musical ideas.

- 30.27 Cues for listening comprehension.** In speaking, we often signal a shift to the next step in the discussion by saying things like “Let’s talk about . . .” or “Finally, . . .” These are “discourse-signaling cues.” Do such cues help listeners better comprehend a second language? One study of this question involved 80 Korean learners of English as a foreign language.¹⁶ Half of the 80 learners listened to a lecture in English with discourse-signaling cues, and the other half heard the same lecture without cues. After the lecture, half of each group were asked to summarize information from the lecture and the other half simply to recall information. Here are the group means and two-way ANOVA table for one response variable, a measure of comprehension of high-level information:



		Lecture	
		Cues	No Cues
Task	Summary	23	17
	Recall	19	13

Source	df	Sum of Squares	<i>F</i>	<i>P</i>
Lecture	1	649.80	16.582	0.000
Task	1	281.25	7.177	0.009
Interaction	1	0.20	0.005	0.943
Error	76	2978.30		

Use this information to comment on the impact of discourse-signaling cues and the type of task performed. (Assume that the data satisfy the conditions for ANOVA.)

- 30.28 Fertilization of bromeliads.** Does the type of fertilization have a significant effect on leaf development for bromeliads? Researchers compared leaf production and death over a seven-month period for a control

group (C), a group fertilized only with nitrogen (N), a group fertilized only with phosphorus (P), and a group that received both nitrogen and phosphorus (NP).¹⁷ In the nitrogen and phosphorus columns in the data file, 0 indicates no fertilization and 1 indicates fertilization. The “new” and “dead” columns give the total number of new or dead leaves produced over the seven months following fertilization. “Change” gives the difference “new” minus “dead.” Give the two-way ANOVA table for the number of new leaves. What do you conclude? BROMLIAD

- 30.29 Fertilization of bromeliads, continued.** Repeat the analysis of the previous exercise using number of dead leaves as the response variable.

- 30.30 Fertilization of bromeliads, continued.** Repeat the analysis of Exercise 30.28 using change in the number of leaves as the response variable. How does the change in response affect your conclusions? How would you summarize the results in these three exercises?

- 30.31 ANOVA with one observation per treatment (optional).** If a two-way layout has just one observation for each treatment, the mean square for error (MSE) is 0 because there is no variation within each treatment group. The usual two-way ANOVA *F* tests can’t be done because they have MSE in their denominators. But if we are willing to assume that there is no interaction between the two factors, the interaction mean square can be used as the denominator of *F* statistics for testing the two main effects. Software usually allows you to choose “no interaction.” Here is an example.

Does the addition of sodium chloride change brick quality? One measure of quality is “resistance anisotropy,” which combines the results of ultrasound and mechanical tests. Researchers tested two types of bricks, one without sodium chloride (MPW) and one with sodium chloride (MPS), at three firing temperatures (850, 1000, and 1100 degrees Celsius). Here are the data for one brick per treatment:¹⁸ BRICKS

Brick Type	Temperature (°C)	Resistance
MPW	850	3.29
MPW	1000	5.78
MPW	1100	4.84
MPS	850	2.57
MPS	1000	3.90
MPS	1100	2.57

- (a) Make a plot of the group means (which are the same as the individual observations). Is the interaction between type of brick and firing temperature small? If so, it may be reasonable to assume that in the population there is no interaction.

- (b) The researchers did fit a two-way ANOVA model without interaction. Use statistical software to obtain this two-way ANOVA table. What do you conclude about the main effects?

30.32 Packaging and perceived healthiness. Can the font and color used in a packaged product affect our perception of the product's healthiness? It was hypothesized that use of a natural font, which looks more handwritten and tends to be more slanted and curved, might lead to a different perception of product healthiness than an unnatural font. Two fonts, Impact and Sketchflow Print, were used. These fonts were shown in a previous study to differ in their perceived naturalness, but otherwise were rated similarly on factors such as readability and likeability. The two package colors used were red and green. Images of four identi-



cal packages differing only in the font and color were presented to the subjects. Participants read statements such as "This product is healthy"/natural/whole-some/organic and they rated how much they agreed with the statement on a seven-point scale, with 1 indicating "strongly agree" and 7 indicating "strongly disagree." Each subject's responses were combined to create a perceived healthiness score.¹⁹ The researcher had 252 students available to serve as subjects and randomly assigned 63 to each treatment, or combination, of color and font. The data file records the perceived healthiness score for each subject. Carry out a complete analysis of the effects of color and font on perceived healthiness, using Example 30.6 to guide your work.



COOKIES

EXPLORING THE WEB

30.1 Find a two-way ANOVA. Find an example of a two-way ANOVA on the web. The *Journal of the American Medical Association* (jama.ama-assn.org), *Science Magazine* (www.sciencemag.org), the *Canadian Medical Association Journal* (www.cmaj.ca), and the *Journal of Marketing Research* (www.atypon-link.com/AMA/loi/jmkr) are possible sources. To help locate an article, look through the abstracts of articles. If you are having difficulty, consider study 1 in *Psychological Science*, 19 (2008), pp. 268–273. Once you find a suitable article, read the article and then briefly describe the study (including the two factors and the number of levels of each factor) and its conclusions. If the means are given, plot them and discuss the interaction. Whether or not the study is balanced, the test for interaction is interpreted in the same way. If *P*-values are reported, be sure to discuss them in your summary. Also, be sure to give the reference (either the web link or the journal, issue, year, title of the paper, authors, and page numbers).

Notes and Data Sources

1. We thank Charles Cannon of Duke University for providing the data. The study report is C. H. Cannon, D. R. Peart, and M. Leighton, "Tree species diversity in commercially logged Bornean rainforest," *Science*, 281 (1998), pp. 1366–1367.
2. Brad J. Bushman, "Violence and sex in television programs do not sell products in advertisements," *Psychological Science*, 16 (2005), pp. 702–707.
3. Alex B. Van Zant and Don A. Moore, "Leaders' use of moral justifications increases policy support," *Psychological Science*, 26 (2015), pp. 934–943.
4. In some cases, we think of the blocks in a block design as a sample from some larger population of potential blocks. For example, a clinical trial may compare several medical therapies (the factor) using patients at several medical centers (the blocks). We are not interested in these particular medical centers; they simply represent all hospitals at which the therapies might be used. Such blocks are called "random." ANOVA with one or more random factors is an advanced topic. In other cases, the specific blocks are of interest in their own right. For example, a clinical trial might compare several therapies (the factor) on both female and male patients (the blocks). We randomly assign therapies among the women and separately among the men, a randomized block design. We are interested in the effect of gender on response to the therapies. Such blocks are called "fixed." In this chapter, we allow randomized block

designs with fixed blocks, analyzing the data by treating the blocks as the second factor in two-way ANOVA.

5. Victoria L. Brescoll and Eric L. Uhlmann, "Can an angry woman get ahead? Status conferral, gender and expression of emotion in the workplace," *Psychological Science*, 19 (2008), pp. 268–273. The description and data are based on study 1 in this article.
6. Orit E. Hetzroni, "The effects of active versus passive computer-assisted instruction on the acquisition, retention, and generalization of Blissymbols while using elements for teaching compounds," PhD thesis, Purdue University, 1995.
7. Data courtesy of David LeBauer, University of California, Irvine. Provided by Brigitte Baldi.
8. Based on Tina Poon and Bianca Grohmann, "Spatial density and ambient scent: Effects on consumer anxiety," *American Journal of Business*, 29 (2014), pp. 76–94. We have used the summary statistics and total sample size in the paper. Although the original sample design had a slight imbalance, the ANOVA table in our example agrees closely with the original paper.
9. We thank Andy Niemiec and Robbie Molden for data from a summer science project at Kenyon College. Provided by Brad Hartlaub.
10. We thank Harihuko Itagaki and Andrew Veerde for data from a summer science project at Kenyon College. Provided by Brad Hartlaub.
11. Data from the EESEE story "Stepping Up Your Heart Rate."
12. Christopher Eric Mowen, "Use of glusofinate in glusofinate resistant corn hybrids," MS thesis, Purdue University, 1999.
13. We thank Coach Brian Luthy for the data. Provided by Brad Hartlaub.
14. Susanne Wiking et al., "Sex differences in furniture assembly performance: An experimental study," *Applied Cognitive Psychology*, 30 (2016), pp. 226–233. We would like to thank the authors for providing the data.
15. John Kratus, "Effect of available tonality and pitch options on children's compositional processes and products," *Journal of Research in Music Education*, 49 (2001), pp. 294–306.
16. Euen Hyuk Jung, "The role of discourse signaling cues in second language listening comprehension," *Modern Language Journal*, 87 (2003), pp. 562–577.
17. We thank Jacqueline Ngai for providing data from Jacqueline T. Ngai and Diane S. Srivastava, "Predators accelerate nutrient cycling in a bromeliad ecosystem," *Science*, 314 (2006), p. 963. Counts for the last control plant are simulated because a monkey enjoyed snacking on this plant before the researchers were able to obtain all the counts.
18. G. Cultrone et al., "Ultrasound and mechanical tests combined with ANOVA to evaluate brick quality," *Ceramics International*, 27 (2001), pp. 401–406.
19. Alysha Fligner and Xiaoyan Deng, "The effect of font naturalness on perceived healthiness of food products," Senior thesis, The Ohio State University.