CHAPTER 13

Two-Factor Experiments: Balanced, Completely Randomized, Factorial Designs

IN THIS CHAPTER

Balanced design
Completely randomized design
Factorial design
Two-factor experiments
Two-way analysis of variance
Replication

There are many situations in which we are interested in the effects of more than a single factor on a response variable. Imagine, for instance, that you run a weight loss clinic. For any given client, you have a choice of several diets, as well as a choice of several exercise regimens. You realize that different diets might result in different average weight loss among your clients. Also, average weight loss may depend on exercise program. In addition, you suspect that the effectiveness of a given diet may depend on the exercise regimen being followed by a client. If the relative effects of the different diets depend on the particular exercise regimen (or, equivalently, the relative effectiveness of the different exercise programs depends on diet), we say there is an *interaction effect* of diet and exercise on weight loss.

> We say two factors have an **interaction effect** on response if the relative effects of one factor on response depend on the level of the other factor.

Suppose a number of your clients have consented to participate in an experiment that you design. To compare the several diets with respect to weight loss, you could set up a single-factor experiment, assigning a number of clients to each of the diets. You would try to keep extraneous sources of variation to a minimum in order to assess the effects of the diets on weight loss. One possible extraneous factor might be exercise program; you would want all participants in this experiment to maintain the same exercise regimen. What are some other possible extraneous factors?

Similarly, you could carry out a single-factor experiment to compare the several exercise programs with respect to weight loss. A possible extraneous factor in this experiment is diet; you would want all clients participating in this experiment to be on the same diet. Other possible extraneous factors would include those you listed above.

Now you have done two separate single-factor experiments—one to compare diets and one to compare exercise programs. You kept exercise program constant in the diet experiment and you kept diet constant in the exercise experiment. Therefore, you are not able to assess the extent of any interaction effect of diet and exercise on average weight loss.

A way around this problem is to include diet and exercise in the same experiment; we call this a *two-factor experiment.*

> In a **two-factor experiment,** we are interested in the effects of two factors, and possible interaction effects of the two factors, on a response variable.

If you include all combinations of diets and exercise programs in your experiment, you have a *factorial design.*

> In a **factorial design,** all combinations of levels of the factors are included in the experiment.

Your experiment is *balanced* if you assign the same number of clients to each combination of diet and exercise.

> An experimental design is **balanced** if there is the same number of observations for each combination of factors.

When you divide the clients among the diet/exercise combinations, you should use *random assignment* so that no bias, whether subconscious or not, affects the way clients are assigned. We also hope that random assignment will tend to balance any extraneous factors that might affect our experimental results. When we randomly assign experimental units to combinations of the factors (such as diet and exercise), we have a *completely randomized design.* (In the diet/exercise example, the experimental units are the clients participating in the experiment.)

> An experiment is **completely randomized** if experimental units are randomly assigned to combinations of factors.

To look for a possible interaction effect of diet and exercise on average weight loss, you need to assign at least two clients to each combination of diet and exercise. Then you have *replication.*

> We have **replication** in an experiment if we have more than one observation per factor combination.

If you use a balanced, completely randomized, factorial design with replication you will be able to compare diets, compare exercise programs, as well as assess possible interaction effects of diet and exercise on average weight loss. Each of these three comparisons will be independent of the other two. With a well-designed two-factor experiment, you can economically obtain more information than would be available from separate single-factor experiments.

A randomized block experiment (k treatments, b blocks) is a special type of two-factor experiment. One factor is *treatment,* with k *levels,* corresponding to the k treatments. The other factor is *block,* with b *levels,* corresponding to the b blocks. In this chapter we are concerned with two-factor experiments other than randomized block designs.

Suppose the first factor, factor A, has I levels and the second factor, factor B, has J levels. If we include each of the $I \times J$ combinations of factors A and B in the experiment, then we have a *factorial design.* To allow for the possibility of an interaction effect in our analysis, we must have more than one observation per combination of factors. Then we have *replication.* If we have the same number of observations per combination, we have a *balanced* design.

Suppose we have n observations for each of the $I \times J$ factor combinations, with n greater than or equal to 2 (that is, we have n replications). Then our total sample size is $N = I \times J \times n$. If the N experimental units are randomly assigned to the factor combinations, we have a *completely randomized design.* In this chapter, we discuss the classical, parametric, analysis of *balanced, completely randomized, factorial experiments with two factors.*

We will assume that the I levels of factor A are the specific levels of interest to us. We say factor A has *fixed effects.* Similarly, the J levels of factor B are the specific levels of interest to us, so factor B also has fixed effects. (In the diet and exercise example, you are interested in the specific diets and the specific exercise programs included in the experiment.) In contrast, the blocks in a randomized block design are *random effects.* We generally are not interested in the specific levels of a random effect variable, whether people, dogs,

or whatever; those included in the experiment are representative of a larger group of subjects of interest to us. We will not consider experiments with random effects (other than the simple randomized block design we have already seen). For discussion of such experiments, see a book on experimental design such as the one by Kirk (1982).

In Section 13-1, we consider the classical, parametric, analysis of two-factor experiments that have balanced, completely randomized, factorial designs. This analysis is called *two-way analysis of variance* or *two-factor analysis of variance.*

> **Two-way** (or **two-factor**) **analysis of variance** is the classical, parametric, approach to analyzing the results of a two-factor experiment that has a balanced, completely randomized, factorial design.

We consider a special case in Section 13-2, when each factor has two levels. Such experiments are especially useful in exploratory studies.

13-1 Two-Factor Analysis of Variance

Let's first look at an example of a balanced, completely randomized, factorial experiment.

EXAMPLE 13-1

How well do different laboratories agree in measuring niacin in bread? Does the extent of agreement depend on the amount of niacin in the bread? Researchers designed a study to answer these questions. They divided samples of bread into three groups. The bread in the first group was not enriched with niacin. The bread in the second group was enriched with 2 milligrams of niacin per 100 grams of bread. The bread in the third group was enriched with 4 milligrams of niacin per 100 grams of bread. The experimenters sent samples to each of six laboratories, where laboratory workers divided a sample into three subsamples. They measured niacin in the sample on three different days (one subsample each day), so there were three replications for each combination of laboratory and niacin level. The measurements of niacin (in milligrams per 100 grams) are shown below (a portion of the data in Rice, 1988, pages 429–430; from Campbell and Pelletier, 1962).

Niacin enrichment	Laboratory					
	a	b	c	d	e	f
0	3.40	3.80	3.66	4.37	4.20	3.76
	3.63	3.80	3.92	3.86	3.60	3.68
	3.52	3.90	4.22	4.46	4.20	3.80
2	5.00	5.30	5.68	6.53	5.80	6.06
	5.27	5.60	5.47	5.85	5.70	5.60
	5.39	5.80	5.84	6.38	5.90	6.05
4	6.54	7.10	7.30	8.32	7.70	7.60
	7.46	7.60	6.40	7.34	7.10	7.50
	6.84	8.00	7.60	8.12	7.20	7.67

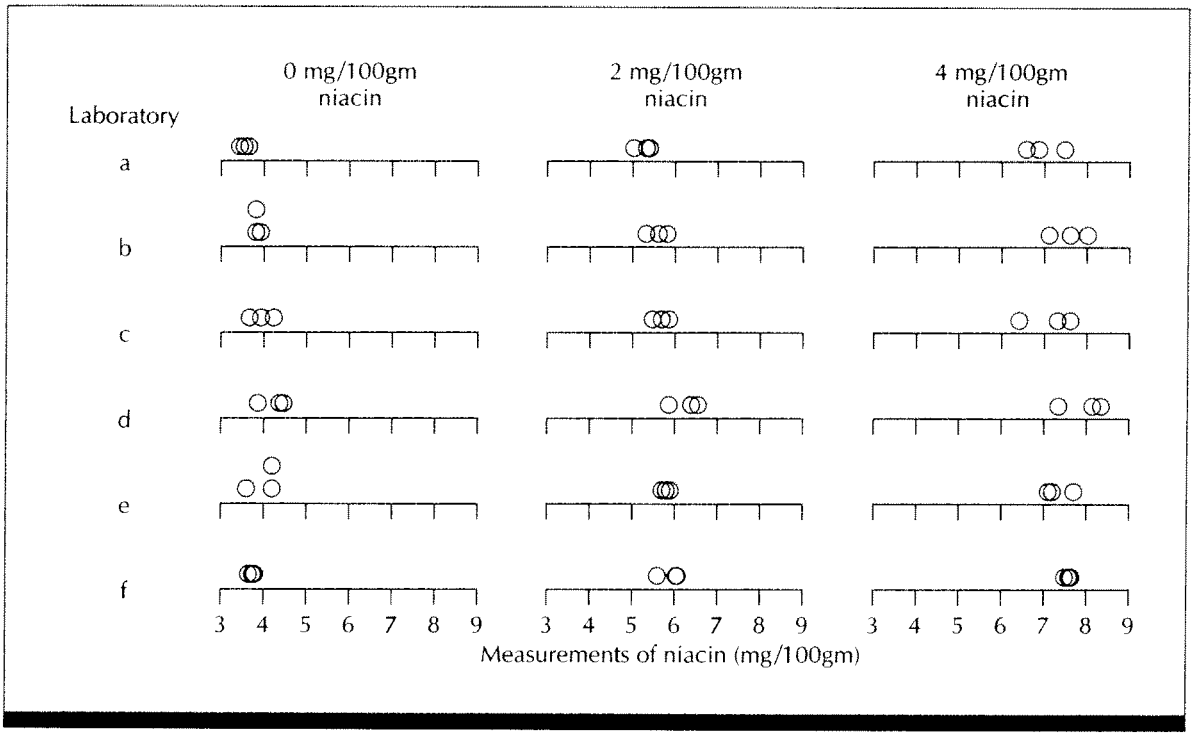

FIGURE 13-1 Plots of niacin measurements in Example 13-1, by laboratory and niacin enrichment level

Do you have suggestions for these investigators? How would you decide which samples go to which laboratories? Should we be concerned about the order in which the samples are analyzed in each laboratory? Would you worry about the time of day, the technician doing the measurements, whether the technician has knowledge of the niacin enrichment level of the sample? Give detailed instructions (called a protocol) for carrying out this experiment.

Plots of the niacin measurements are shown in Figure 13-1 by laboratory and niacin enrichment level. Does it look like the laboratory determinations of niacin depend on the level of niacin enrichment? Do there appear to be differences among laboratories? How does the spread or variation among values compare across the 18 plots?

We are also interested in possible interaction effects on measurements. Do differences among laboratories depend on the level of niacin enrichment? Alternatively, do the relative effects of niacin enrichment level depend on the laboratory making the measurements? One way to answer this question is through a plot such as shown in Figure 13-2. This figure shows a scatterplot of average niacin measurement versus enrichment level for each of the six laboratories.

The points for a single laboratory in Figure 13-2 are connected, creating a *profile* for that laboratory. The slope of each profile clearly illustrates the

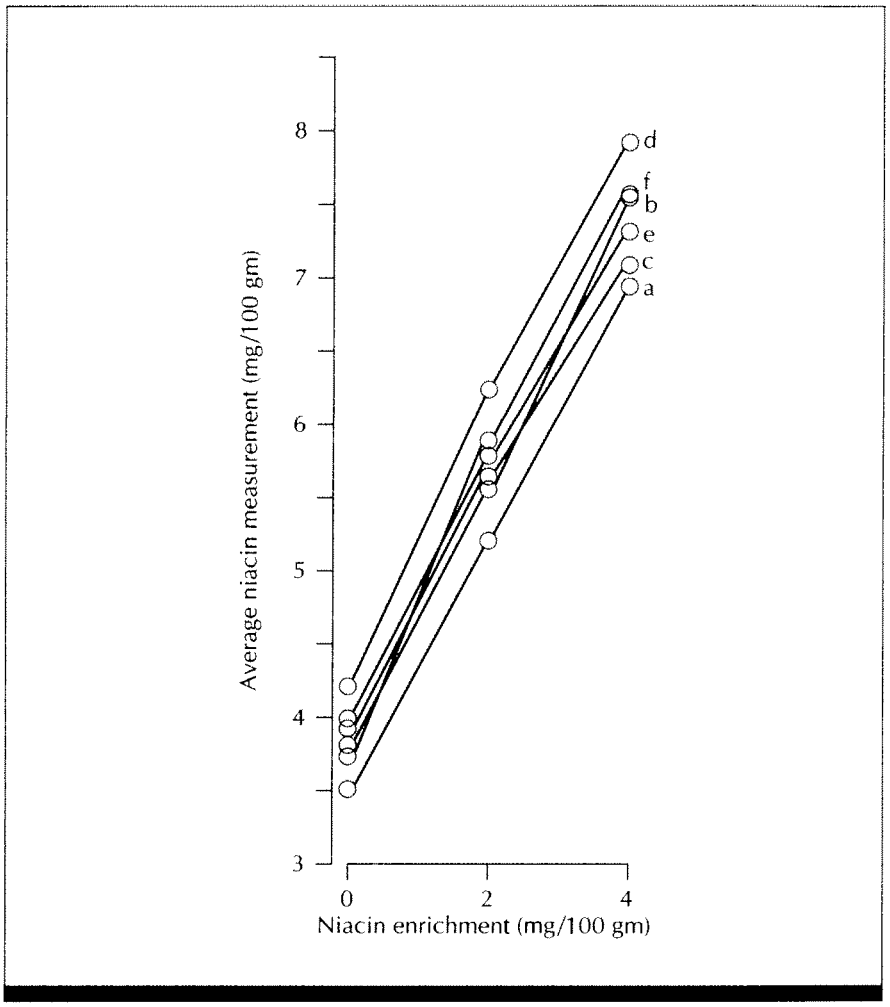

FIGURE 13-2 Scatterplot of average niacin measurement versus niacin enrichment level for each laboratory. The points for a single laboratory are connected.

differences in average measurements for the three niacin enrichment levels. The profiles for the six laboratories do not all coincide (they are not all on top of each other), so there may be differences among laboratories. However, the profiles are parallel: the relative effects of niacin enrichment level seem to be about the same for all six laboratories. That is, there seems to be no interaction effects of laboratory and niacin enrichment level on the niacin measurements.

A formal analysis of this experiment tests hypotheses about differences among laboratories, differences among niacin enrichment levels, and interaction effects of laboratory and niacin enrichment level on average niacin determination. Before outlining the method of analysis, we need some notation.

Two-Factor Analysis of Variance for a Balanced Factorial Design

Suppose factor A has I levels and factor B has J levels. There are n replications per factor combination, with n greater than or equal to 2. Let Y_{ijk} denote the observation for the kth replication at level i of factor A and level j of factor B. $\bar{Y}_{ij}$ represents the average of the n experimental units receiving level i of factor A and level j of factor B. $\bar{A}_i$ denotes the average of the $J \times n$ observations at level i of factor A. Similarly, $\bar{B}_j$ denotes the average of the $I \times n$ observations at level j of factor B. Finally, $\bar{Y}$ represents the average of all $N = I \times J \times n$ observations. The notation for these averages is summarized in Table 13-1, along with four estimates of variation.

The *residual mean square* s_r^2 is a measure of random variation among the observations in the experiment. The *mean square for factor A*, s_A^2, is a measure of random variation plus differences in average response at different levels of factor A. The *mean square for factor B*, s_B^2, measures random variation plus differences in average response at different levels of factor B. The *interaction mean square*, s_{AB}^2, estimates random variation plus nonadditive effects of the two factors on average response.

If the I levels of factor A all have the same average effect on response, then s_A^2 and s_r^2 both estimate random variation and should be similar in magnitude. If the levels of factor A do not all have the same average effect on response, then we expect s_A^2 to be larger than s_r^2. This forms the basis for testing hypotheses about differences on average response among the levels of factor A.

TABLE 13-1 Notation for the averages in a two-factor experiment

	Factor B				
Factor A	1	2	$\cdots$	J	Factor A averages
1	$\bar{Y}_{11}$	$\bar{Y}_{12}$	$\cdots$	$\bar{Y}_{1J}$	$\bar{A}_1$
2	$\bar{Y}_{21}$	$\bar{Y}_{22}$	$\cdots$	$\bar{Y}_{2J}$	$\bar{A}_2$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
I	$\bar{Y}_{I1}$	$\bar{Y}_{I2}$	$\cdots$	$\bar{Y}_{IJ}$	$\bar{A}_I$
Factor B averages	$\bar{B}_1$	$\bar{B}_2$	$\cdots$	$\bar{B}_J$	$\bar{Y}$

Variance estimates or mean squares:

$$s_A^2 = \text{Mean square for factor A} = \frac{J \times n}{I - 1} \sum_{i=1}^{I} (\bar{A}_i - \bar{Y})^2$$

$$s_B^2 = \text{Mean square for factor B} = \frac{I \times n}{J - 1} \sum_{j=1}^{J} (\bar{B}_j - \bar{Y})^2$$

$$s_{AB}^2 = \text{Mean square for AB interaction}$$

$$= \frac{n}{(I - 1)(J - 1)} \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{Y}_{ij} - \bar{A}_i - \bar{B}_j + \bar{Y})^2$$

$$s_r^2 = \text{Residual mean square} = \frac{1}{I \times J \times (n - 1)} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{ij})^2$$

Similarly, if the J levels of factor B all have the same average effect on response, then s_B^2 and s_r^2 both estimate random variation, so we expect them to be similar in magnitude. On the other hand, if the levels of factor B do not all have the same average effect on response, then we expect s_B^2 to be larger than s_r^2. We use this fact to test for differences among levels of factor B on average response.

If the relative effects of the levels of factor A are the same for all levels of factor B (or, equivalently, the relative effects of the levels of factor B are the same for all levels of factor A), then s_{AB}^2 and s_r^2 both estimate random variation and should be similar in magnitude. If the relative effects of the levels of factor A depend on the level of factor B (or the relative effects of levels of factor B depend on the level of factor A), then we expect s_{AB}^2 to be larger than s_r^2. This provides the rationale for our test of an interaction effect of factors A and B on average response.

With these ideas in mind, we can outline the significance level approach to the parametric analysis of a two-factor experiment with a balanced, completely randomized, factorial design. After outlining the approach, we will apply it to Example 13-1.

The significance level approach to two-factor analysis of variance for a balanced factorial design

1. We want to test three sets of hypotheses. One test is about the effect of factor A on the response:

   H_0(A): The I levels of factor A have the same average effect on response.
   H_a(A): The average effect on response is not the same for all I levels of factor A.

   Another test is about the effect of factor B on response:

   H_0(B): The J levels of factor B have the same average effect on response.
   H_a(B): The average effect on response is not the same for all J levels of factor B.

   In addition, we want to test the null hypothesis that the levels of factor A have the same relative effects on response within each level of factor B. This is the same as saying that the levels of factor B have the same relative effects on response within each level of factor A. Then we say there is no interaction effect of factors A and B on response.

   H_0(AB): The relative effects of one factor do not depend on the level of the other factor.
   H_a(AB): The relative effects of one factor do depend on the level of the other factor.

2. To test the hypotheses about the factor A effects, we use the test statistic

$$\text{Test statistic(A)} = \frac{s_A^2}{s_r^2}$$

To test the hypotheses about the factor B effects, we use the test statistic

$$\text{Test statistic(B)} = \frac{s_B^2}{s_r^2}$$

To test hypotheses about the interaction effect, we use the test statistic

$$\text{Test statistic(AB)} = \frac{s_{AB}^2}{s_r^2}$$

3. We assume that the $N = I \times J \times n$ observations are all independent, from Gaussian distributions. These distributions all have the same variance σ^2. The means may differ, depending on the combinations of the two factors.

   Under the null hypothesis that the levels of factor A have the same average effect, test statistic(A) has the F distribution with $I - 1$ numerator degrees of freedom and $IJ(n - 1)$ denominator degrees of freedom.

   Under the null hypothesis that the levels of factor B have the same average effect, test statistic(B) has the F distribution with $J - 1$ numerator degrees of freedom and $IJ(n - 1)$ denominator degrees of freedom.

   Under the null hypothesis of no interaction effect on response, test statistic(AB) has the F distribution with $(I - 1)(J - 1)$ numerator degrees of freedom and $IJ(n - 1)$ denominator degrees of freedom.

   For each of the three sets of hypotheses, small values of the test statistic, near 1, are consistent with the corresponding null hypothesis, while large values are inconsistent with that null hypothesis.
4. Select a significance level for each test.
5. Let α denote the significance level for one of the three sets of hypotheses. Find the number c from Table D such that $P(F \leq c) = 1 - \alpha$. Here, F denotes a random variable having the F distribution with degrees of freedom corresponding to the hypotheses being tested. The acceptance region is $[0, c)$; the rejection region is $[c, \infty)$.
6. For each set of hypotheses, the decision rule has the form:

   If test statistic $< c$, say the results are consistent with the null hypothesis.
   If test statistic $\geq c$, say the results are inconsistent with the null hypothesis.
7. Carry out an experiment that satisfies the assumptions in step 3. Calculate the test statistics in step 2. Use appropriate decision rules to decide whether there seem to be effects of the two factors and/or interaction effects on response. Draw conclusions based on the experimental results.

EXAMPLE 13-1 *(continued)*

Let's return now to the experiment described in Example 13-1. Call niacin enrichment level factor A and laboratory factor B. We want to test three sets of hypotheses. The first set is about the effect of niacin enrichment level:

H_0(A): All three niacin enrichment levels have the same average effect on niacin determination.

H_a(A): The three niacin enrichment levels do not all have the same average effect on niacin determination.

The second set of hypotheses is about differences among laboratories:

H_0(B): The six laboratories all have the same average effect on niacin determination.

H_a(B): The six laboratories do not all have the same average effect on niacin determination.

The third set is about interaction effects:

H_0(AB): The relative effects of niacin enrichment on the measurements are the same for all six laboratories. (Equivalently, relative differences among laboratories are the same for all three niacin enrichment levels.)

H_a(AB): The relative effects of niacin enrichment on the measurements are not the same for all six laboratories. (Relative differences among laboratories are not the same for all three niacin enrichment levels.)

For a parametric analysis of this experiment, we assume that the 54 observations are all independent, from Gaussian distributions having the same variance. From the plots of the observations in Figure 13-1, the equal-variance assumption seems reasonable. We cannot assess the independence assumption without more information on how the experiment was conducted. What suggestions would you have for carrying out the experiment, in order to ensure independence? What other suggestions would you make, to control extraneous sources of variation and make tests of hypotheses valid?

To assess the Gaussian assumption, we can look at residuals. Recall that a *residual* is the difference between an observation and a summary, estimate, or predicted value of the observation. To calculate a residual in a balanced two-way factorial design, we subtract from each observation Y_{ijk} the estimated expected value $\bar{Y}_{ij}$ of that observation, based on the two-way analysis of variance model. Therefore, a residual has the form $Y_{ijk} - \bar{Y}_{ij}$.

> A **residual** is the difference between an observation and an estimate of its expected value.
>
> In a two-factor analysis of variance model that accounts for possible interaction effects, a residual is the difference $Y_{ijk} - \bar{Y}_{ij}$ between an observation and the average of the observations within the same combination of factors.

The 54 residuals for our example are listed in Table 13-2.

The residuals represent what is left over after we fit a two-way analysis of variance model. If the model is adequate, the residuals should appear to be randomly distributed about a mean of 0. If the Gaussian assumption for the observations is reasonable, the residuals will appear to follow a Gaussian distribution with mean 0. From the histogram of the residuals shown in Figure 13-3, the assumption of Gaussian observations seems reasonable.

TABLE 13-2 Residuals from the two-way analysis of variance model in Example 13-1

Factor A: Niacin enrichment level	Factor B: Laboratory					
	a	b	c	d	e	f
0	−.12	−.03	−.27	.14	.20	.01
	.11	−.03	−.01	−.37	−.40	−.07
	.00	.07	.29	.23	.20	.05
2	−.22	−.27	.02	.28	.00	.16
	.05	.03	−.19	−.40	−.10	−.30
	.17	.23	.18	.13	.10	.15
4	−.41	−.47	.20	.39	.37	.01
	.51	.03	−.70	−.59	−.23	−.09
	−.11	.43	.50	.19	−.13	.08

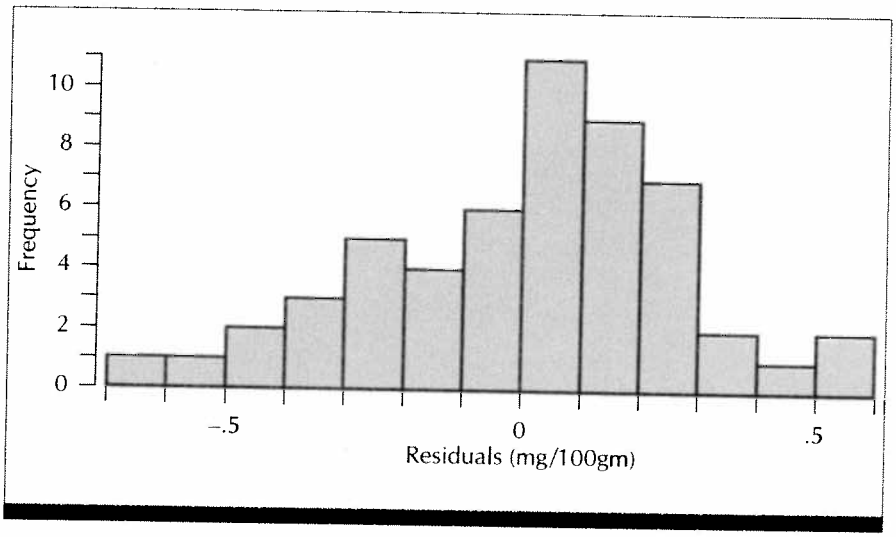

FIGURE 13-3 Histogram of residuals for the two-way analysis of variance model in Example 13-1

Figure 13-4 shows a plot of residuals by niacin enrichment level. The variation in residuals is greater at the highest enrichment level than at the lower levels. The residuals are plotted by laboratory in Figure 13-5. There is a gap in the middle of the plot for laboratory d. Also, the variation in residuals is smaller for laboratory f than for the other laboratories. These plots suggest that not all of our model assumptions may be perfectly met. However, they provide no strong evidence to discourage us from a parametric analysis of the experimental results.

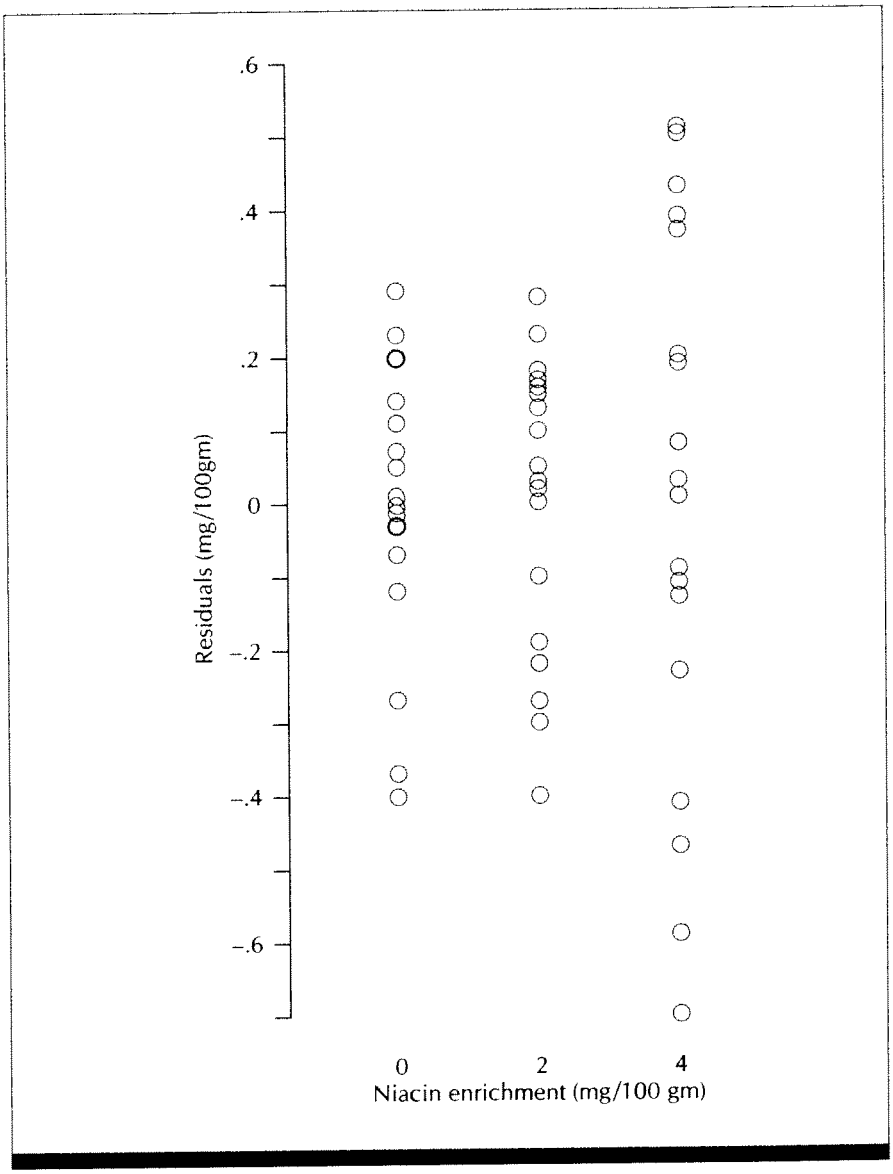

FIGURE 13-4 Plot of residuals by niacin enrichment level in Example 13-1

If all the model assumptions are met, then under the null hypothesis of no difference among niacin enrichment levels with respect to average niacin measurement, test statistic(A) would have the $F(2, 36)$ distribution. Under the null hypothesis of no difference among laboratories in average niacin measurements, test statistic(B) would have the $F(5, 36)$ distribution. Under the null hypothesis of no interaction effects, test statistic(AB) would have the $F(10, 36)$ distribution.

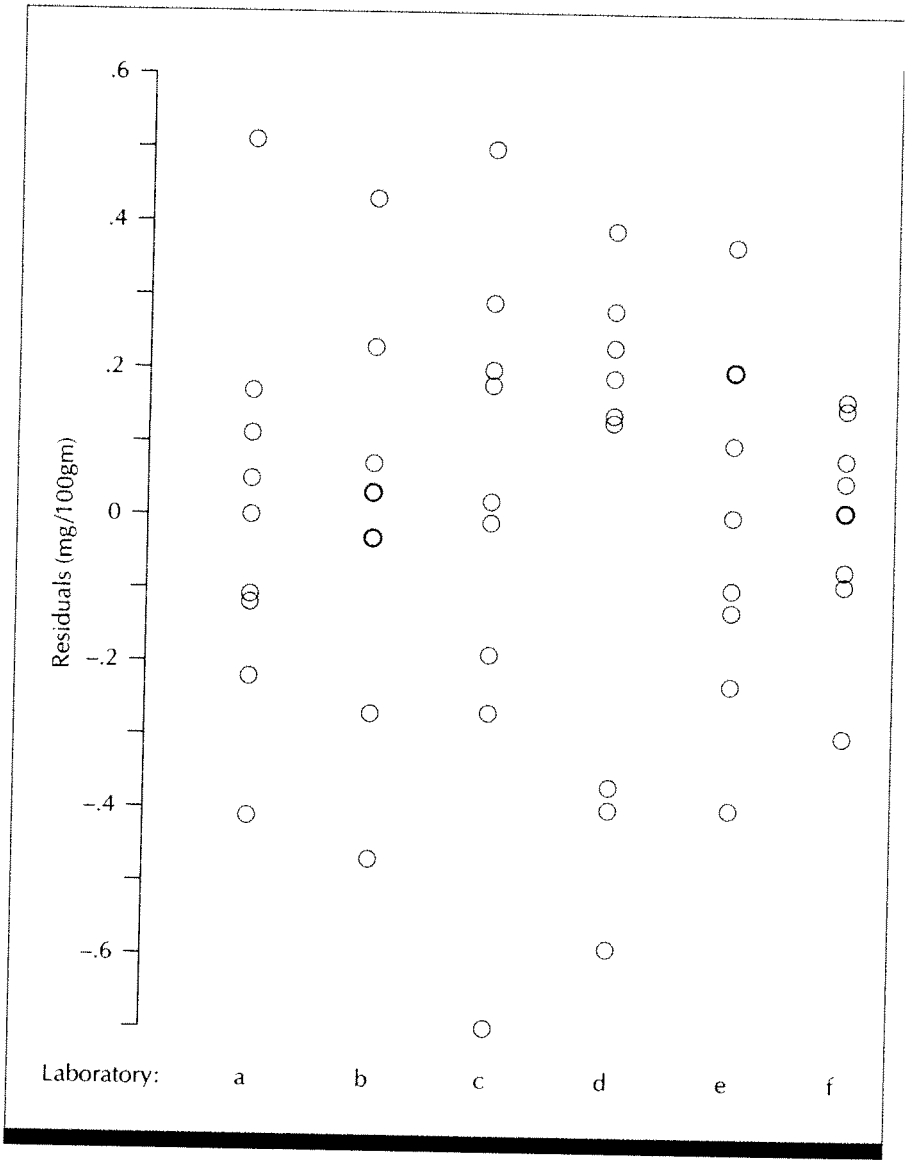

FIGURE 13-5 Plot of residuals by laboratory in Example 13-1

We will use significance level .01 for all three tests of hypotheses. For all three tests of hypotheses, we have 36 denominator degrees of freedom. Since 36 is not a value of d_2 listed in Table D, we will use the value just below 36, $d_2 = 30$. By rounding downward, we get a more conservative test (we are less likely to reject the null hypothesis when it is true).

If F_1 has the $F(2, 30)$ distribution, then $P(F_1 \leq 5.39) = .99$. The acceptance region for the factor A test is [0, 5.39) and the rejection region is

[5.39, ∞). To test for effects of niacin enrichment levels on niacin measurements, the decision rule is:

If test statistic(A) < 5.39, say the results are consistent with the null hypothesis that the levels of niacin enrichment all have the same average effect on niacin determination.

If test statistic(A) ≥ 5.39, say the results are inconsistent with the null hypothesis, suggesting that niacin determinations depend on the level of niacin enrichment.

If F_2 has the $F(5, 30)$ distribution, then $P(F_2 \leq 3.70) = .99$. The acceptance region for the factor B test is [0, 3.70) and the rejection region is [3.70, ∞). To test for laboratory differences, the decision rule is:

If test statistic(B) < 3.70, say the results are consistent with the null hypothesis that the laboratories all have the same average effect on niacin determination.

If test statistic(B) ≥ 3.70, say the results are inconsistent with the null hypothesis, suggesting that the laboratories do not all get the same niacin determinations on average.

If F_3 has the $F(10, 30)$ distribution, then $P(F_3 \leq 2.98) = .99$. The acceptance region for the test of interaction effects is [0, 2.98); the rejection region is [2.98, ∞). To test whether differences among laboratories are the same for all niacin enrichment levels, the decision rule for this interaction effect is:

If test statistic(AB) < 2.98, say the results are consistent with the null hypothesis that the differences among laboratories are the same for all niacin enrichment levels (or the differences among niacin enrichment levels are the same for all laboratories).

If test statistic(AB) ≥ 2.98, say the results are inconsistent with the null hypothesis, suggesting that the differences among laboratories are not the same for all niacin enrichment levels (or the differences among niacin enrichment levels are not the same for all laboratories).

The calculations we need are outlined in Table 13-3.

Test statistic(A) equals 544.7, inconsistent with the null hypothesis that the niacin enrichment levels all have the same average effect on niacin measurements. The *p*-value is less than .01. We see these differences among niacin enrichment levels in the plots of the observations in Figure 13-1 and in the profile plot of average niacin determination versus niacin enrichment level in Figure 13-2.

Test statistic(B) equals 7.6, inconsistent with the null hypothesis that the laboratories all have the same average effect on niacin measurements. The *p*-value is less than .01, so that the differences among laboratories that we saw in the plots are statistically significant. From Figures 13-1 and 13-2 we see that laboratory a tended to get the lowest measurements, laboratory d the highest.

Test statistic(AB) equals .7, consistent with the null hypothesis that the differences in measurements among laboratories are the same for all three niacin enrichment levels (or differences in measurements among niacin en-

TABLE 13-3 Calculations for two-way analysis of variance in Example 13-1. The average niacin measurement $\bar{Y}_{ij}$ is shown below for each combination of niacin enrichment level and laboratory. In parentheses are values of $\bar{Y}_{ij} - \bar{A}_i - \bar{B}_j + \bar{Y}$.

Factor A: Niacin enrichment level	Factor B: Laboratory						Factor A averages
	a	b	c	d	e	f	
0	3.52 (.08)	3.83 (−.04)	3.93 (.15)	4.23 (−.12)	4.00 (.08)	3.75 (−.21)	3.88
2	5.22 (−.06)	5.57 (−.15)	5.66 (.03)	6.25 (.05)	5.80 (.03)	5.90 (.09)	5.73
4	6.95 (−.02)	7.57 (.17)	7.10 (−.21)	7.93 (.05)	7.33 (−.12)	7.59 (.10)	7.41
Factor B averages	5.23	5.66	5.57	6.14	5.71	5.75	5.67

$I = 3 \quad J = 6 \quad n = 3 \quad N = 54 \quad I - 1 = 2 \quad J - 1 = 5 \quad (I - 1)(J - 1) = 10 \quad IJ(n - 1) = 36$

$$s_A^2 = \frac{6 \times 3}{3 - 1}[(3.88 - 5.67)^2 + (5.73 - 5.67)^2 + (7.41 - 5.67)^2] = 56.1$$

$$s_B^2 = \frac{3 \times 3}{6 - 1}[(5.23 - 5.67)^2 + (5.66 - 5.67)^2 + (5.57 - 5.67)^2 + (6.14 - 5.67)^2 + (5.71 - 5.67)^2 + (5.75 - 5.67)^2] = .78$$

$$s_{AB}^2 = \frac{3}{(3 - 1)(6 - 1)}[\text{Sum of the 18 values of } (\bar{Y}_{ij} - \bar{A}_i - \bar{B}_j + \bar{Y})^2] = .07$$

$$s_I^2 = \frac{1}{3 \times 6 \times (3 - 1)}[\text{Sum of the 54 values of } (Y_{ijk} - \bar{Y}_{ij})^2] = .103$$

$$\text{Test statistic(A)} = \frac{56.1}{.103} = 544.7 \qquad \text{Test statistic(B)} = \frac{.78}{.103} = 7.6 \qquad \text{Test statistic(AB)} = \frac{.07}{.103} = .7$$

richment levels are the same for all laboratories). This agrees with the observations we made about the parallel profiles in Figure 13-2.

The test statistic for niacin enrichment levels equals 544.7; this very large test statistic agrees with the obvious differences in measurements across niacin enrichment levels we observe in Figures 13-1 and 13-2. The test statistic for laboratories, 7.6, is much smaller. While statistically significant, it is more difficult for us to make a visual assessment of laboratory differences in niacin determinations by examining Figures 13-1 and 13-2. The lack of an interaction effect on response is very clear in the profile plot in Figure 13-2.

We can summarize the calculations for two-way analysis of variance in a table. A general form of analysis of variance table for a balanced two-way factorial experiment is shown in Table 13-4.

Computer output often shows an additional column for the p-value, at the right of the table. The analysis of variance table for Example 13-1 is given in Table 13-5. The first p-value is listed as 0.0000, meaning that the actual p-value is less than .0001.

You may notice that the values for the mean squares and test statistics in Table 13-5 are not exactly the same as those listed in Table 13-3. Table 13-5

TABLE 13-4 Analysis of variance table for a balanced two-way factorial experiment

Source of variation	Sum of squares	Degrees of freedom	Mean square	Test statistic
Factor A	$Jn\sum_{i=1}^{I}(\bar{A}_i - \bar{Y})^2$	$I - 1$	s_A^2	$\frac{s_A^2}{s_r^2}$
Factor B	$In\sum_{j=1}^{J}(\bar{B}_j - \bar{Y})^2$	$J - 1$	s_B^2	$\frac{s_B^2}{s_r^2}$
Interaction	$n\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{Y}_{ij} - \bar{A}_i - \bar{B}_j + \bar{Y})^2$	$(I - 1)(J - 1)$	s_{AB}^2	$\frac{s_{AB}^2}{s_r^2}$
Residuals	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y}_{ij})^2$	$IJ(n - 1)$	s_r^2	
Total	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y})^2$	$IJn - 1$		

TABLE 13-5 Analysis of variance table for the experiment in Example 13-1

Source of variation	Sum of squares	Degrees of freedom	Mean square	Test statistic	*p*-value
Niacin enrichment	112.494344	2	56.2471721	544.30	0.0000
Laboratories	3.88739403	5	0.777478805	7.52	0.0001
Interaction	.708944299	10	0.07089443	0.69	0.7302
Residual	3.72019973	36	0.103338882		
Total	120.810882	53			

summarizes computer output generated by the student version of the personal computer package Stata®, while the values in Table 13-3 were calculated by hand, using fewer decimal places than the computer uses. The resulting *round-off error* accounts for the differences in calculated results.

Many workers consider the test of interaction effects first in a two-way analysis of variance. The reason: If it looks like there is an interaction effect of the two factors on response, then we can say that each factor appears to have an effect on response. Can you explain why this is so?

In Section 13-2, we discuss a two-factor experiment in which both factors have two levels. Graphs are very easy to interpret in this situation.

13-2 Two-Factor Experiments with Each Factor at Two Levels

We are still interested in a two-way, balanced, completely randomized, factorial design. But now we consider experiments in which each factor has two levels.

Simple plots provide a lot of useful information in this case. Let's look at ar example.

EXAMPLE 13-2 An engineer was interested in the effect of two factors on strength of sheet castings of a polymer. One factor was time in a polymerization bath with 1% catalyst. The other factor was the temperature of the polymerization bath. The engineer tried two times (20 minutes and 60 minutes) and two temperatures (100° and 120°). He treated samples of polymer under each of the four sets of conditions. Then he measured flexural strength (in pounds per square inch, or psi) on samples of sheet castings of the polymer. The results are shown below (Duncan, 1974, page 685 (temperature scale not given); from Gore, 1947).

Time in bath with 1% catalyst	Temperature of polymerization bath	
	100°	120°
20 minutes	9,500	11,300
	10,650	11,750
	9,700	11,600
	9,950	11,650
	10,100	11,700
60 minutes	11,500	10,900
	11,650	11,500
	11,250	11,850
	11,250	11,700
	11,900	11,650

A plot of the observations is shown in Figure 13-6. What does this plot suggest about the effects of time in bath and bath temperature on strength of these sheet castings?

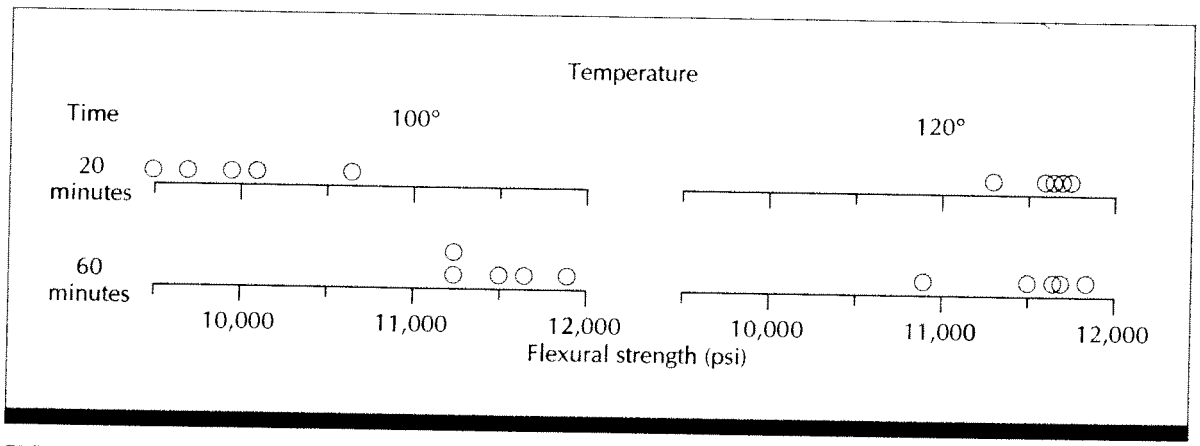

FIGURE 13-6 Plot of the flexural strength measurements in Example 13-2

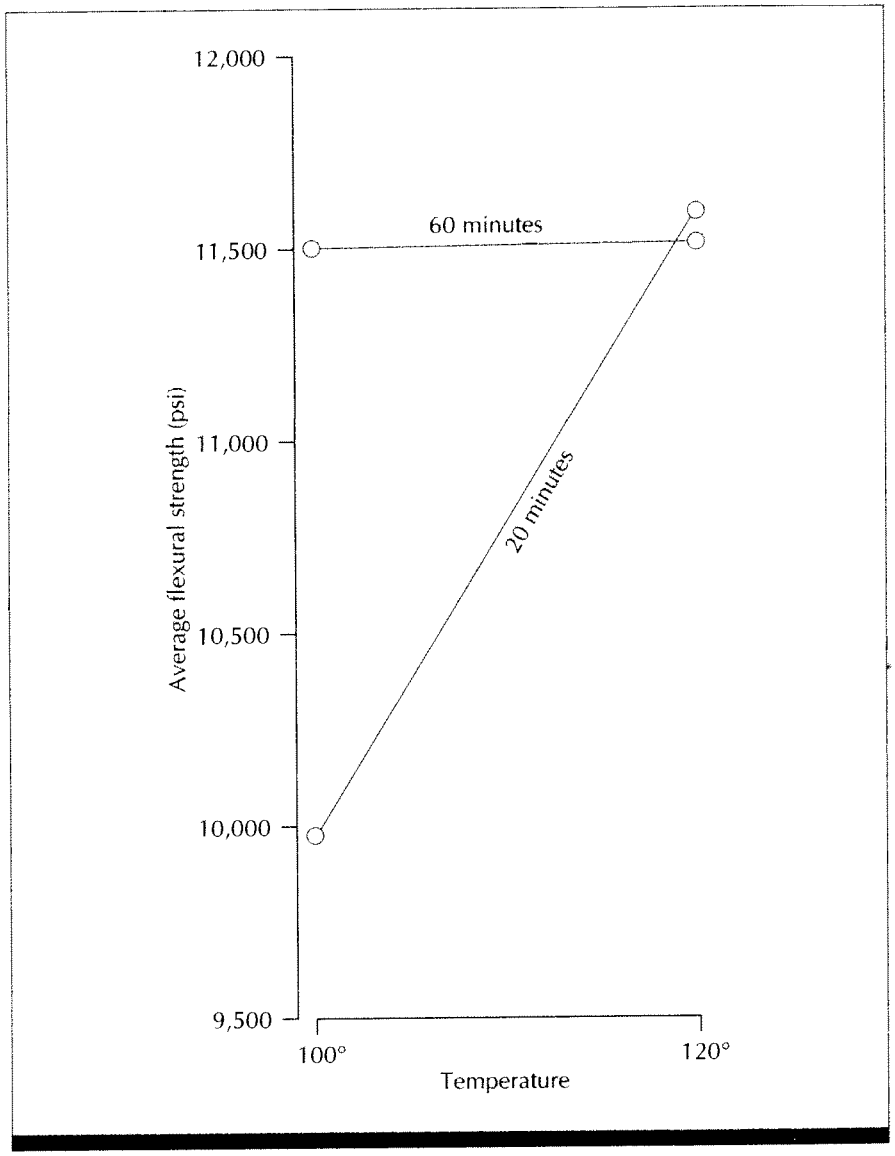

FIGURE 13-7 Plot of average flexural strength versus temperature, for each time. The two points for each time are connected.

Figure 13-7 shows a plot of average flexural strength versus temperature for each of the two bath times. The two points are connected for each time, creating a *profile*. Because the two profiles in Figure 13-7 are not parallel, we can easily see the interaction effect of time and temperature on flexural strength. When time in the bath was 20 minutes, the change in temperature made a big difference in strength of the sheet castings. When time in the bath

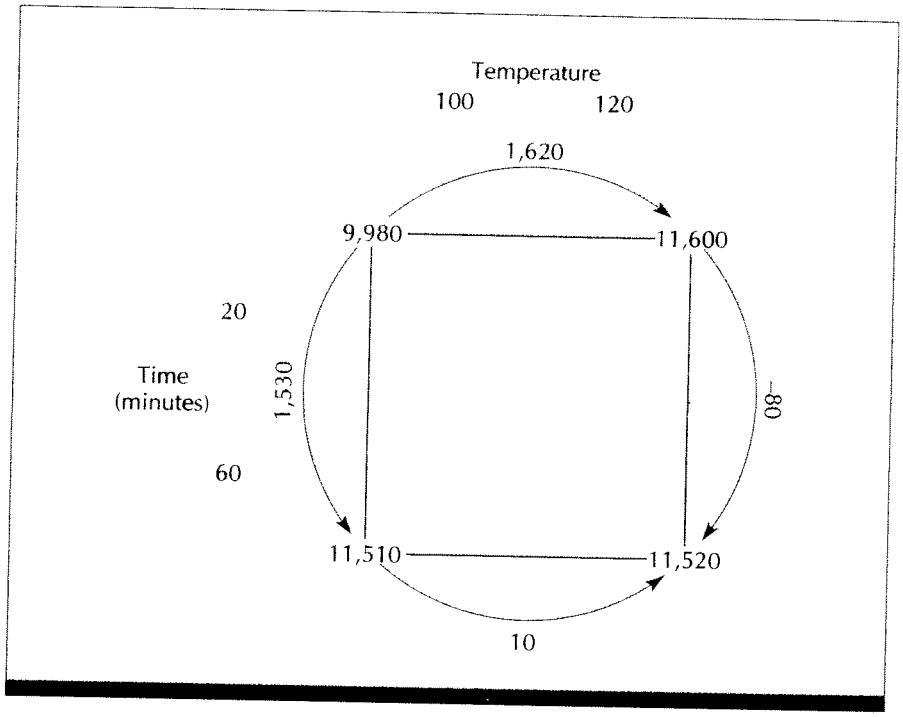

FIGURE 13-8 Plot of average flexural strength (pounds per square inch) under the four sets of conditions in Example 13-2

was 60 minutes, the change in temperature made little difference in strength of the product.

Another useful plot when factors have two levels is shown in Figure 13-8 (Box, Hunter, and Hunter, 1978, Chapter 10). Values of average flexural strength under the four sets of conditions are shown at the corners of the square. We can easily see the direction and extent of the interaction effect of time and temperature on flexural strength. At the low temperature, we see that the longer time in the bath resulted in 1,530 psi greater average flexural strength; at the higher temperature, the longer time in the bath resulted in 80 psi lower average flexural strength. For the shorter time, the higher temperature resulted in 1,620 psi greater average flexural strength; for the longer time, the higher temperature resulted in only 10 psi greater average flexural strength.

For two-way analysis of variance, we assume that we have independent observations from Gaussian distributions with equal variances. From the plot of the observations in Figure 13-6, variation among values does not seem too different across the four samples. We cannot judge the independence assumption without more information on how the experiment was conducted. What suggestions would you make to the experimenter, to ensure independence, reduce the effects of extraneous factors, and allow valid tests of the hypothesis?

To assess the Gaussian assumption, we can look at plots of residuals. A

TABLE 13-6 Residuals for the two-way analysis of variance model in Example 13-2

Time in the bath with 1% catalyst	Temperature of the polymerization bath	
	100°	120°
20 minutes	−480	−300
	670	150
	−280	0
	−30	50
	120	100
60 minutes	−10	−620
	140	−20
	−260	330
	−260	180
	390	130

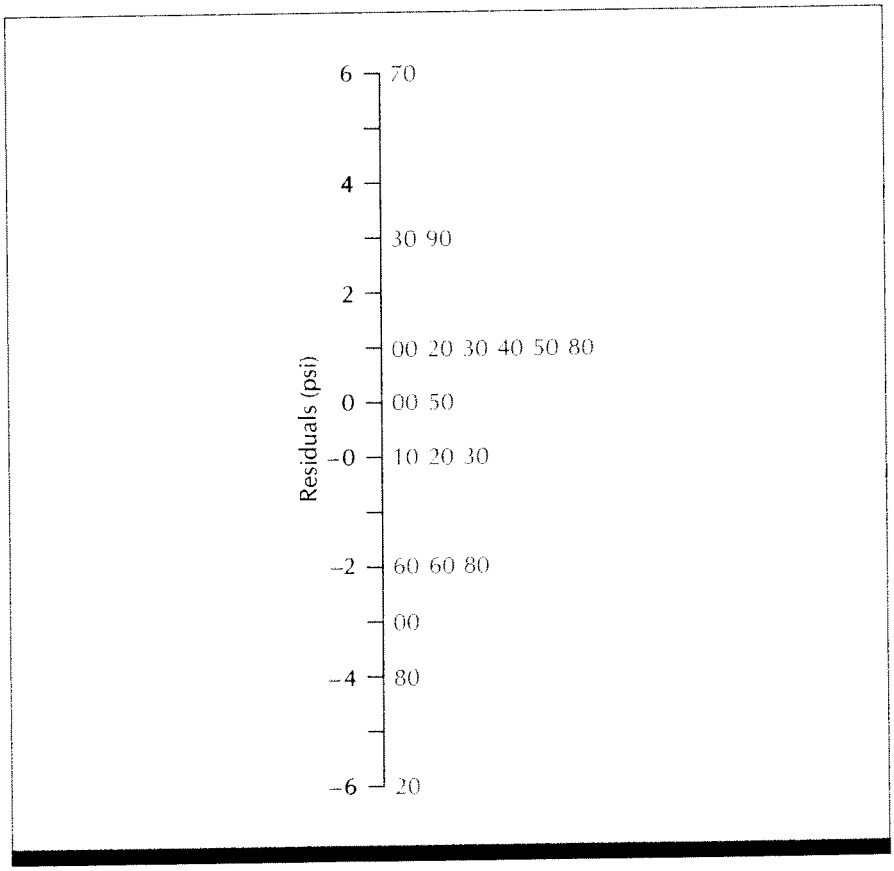

FIGURE 13-9 Stem-and-leaf plot of residuals in Example 13-2. The stems are in hundreds of pounds per square inch. The leaves are in pounds per square inch.

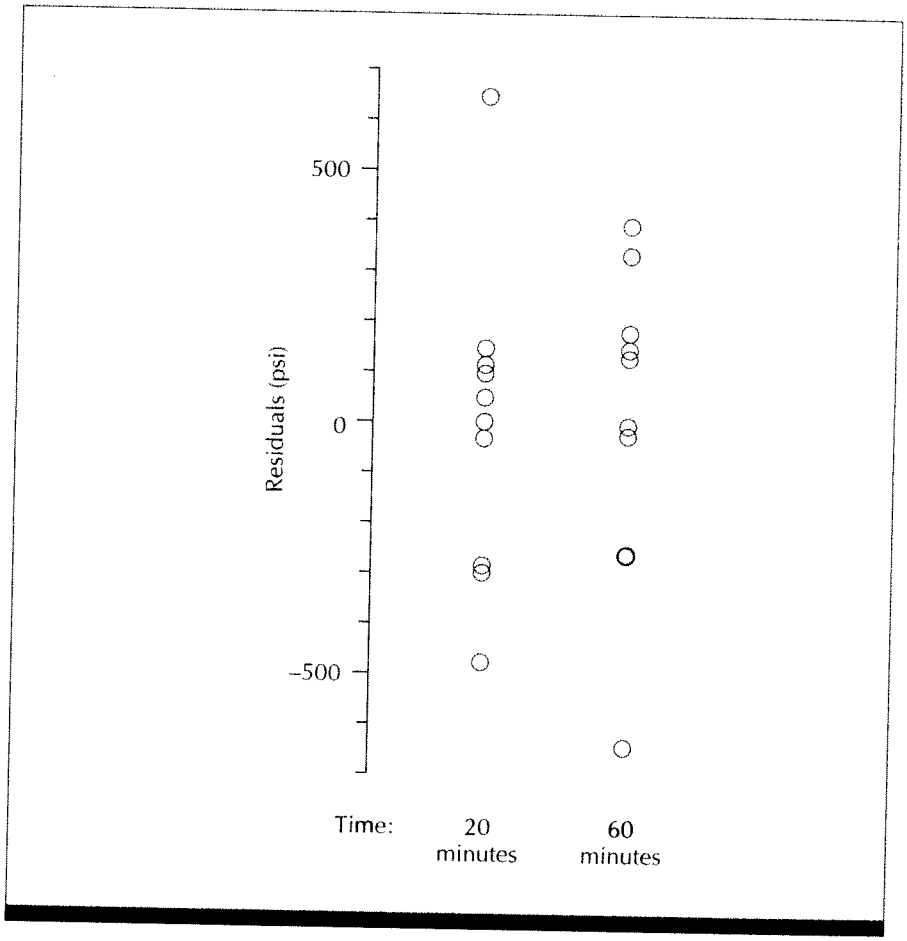

FIGURE 13-10 A plot of residuals by time, Example 13-2

residual for the balanced factorial two-way analysis of variance model is the difference $Y_{ijk} - \bar{Y}_{ij}$ between an observation and the average of all the observations made under the same set of conditions. The residuals for Example 13-2 are shown in Table 13-6.

A stem-and-leaf plot of residuals is shown in Figure 13-9. This plot gives the impression that the assumption of Gaussian observations is reasonable.

Figure 13-10 shows a plot of residuals by time, while Figure 13-11 shows a plot of residuals by temperature. We see that the plot of residuals is shifted up (toward positive values) for the low level of time and temperature. The residuals are shifted down (toward negative values) for the high level of time and temperature. Looking at Table 13-6, we see that the most extreme positive residual is at the low level of time and temperature. The most extreme negative residual is at the high level of time and temperature. Our two-way analysis of variance model cannot account for these anomalies. We will proceed with the analysis, but be cautious in our interpretations. Since all the model assump-

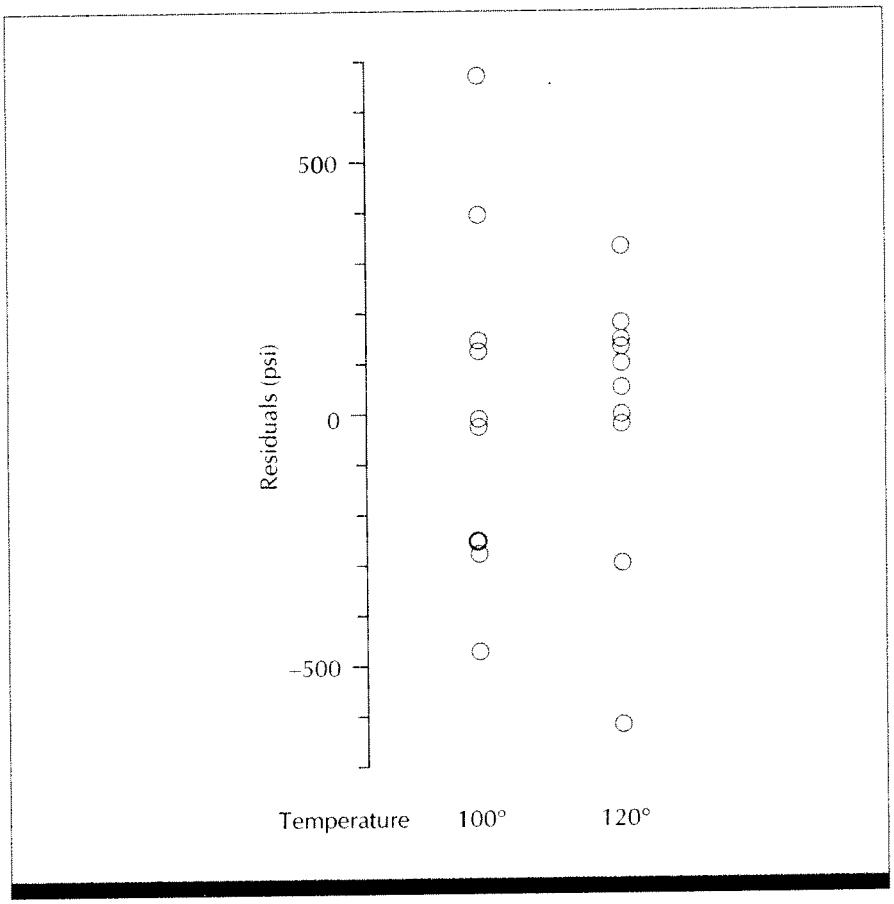

FIGURE 13-11 A plot of residuals by temperature, Example 13-2

TABLE 13-7 Analysis of variance table for the two-factor experiment in Example 13-2

Source of variation	Sum of squares	Degrees of freedom	Mean square	Test statistic	*p*-value
Time	2,628,125	1	2,628,125	24.06	0.0002
Temperature	3,321,125	1	3,321,125	30.40	0.0000
Interaction	3,240,125	1	3,240,125	29.66	0.0001
Residual	1,748,000	16	109,250		
Total	10,937,375	19			

tions may not be met, we will use the formal analysis to show trends, but not interpret *p*-values as though they were exact.

The analysis of variance table from the computer package Stata is shown in Table 13-7. The table corroborates what we saw in the plots. There are statistically significant effects of time and temperature, as well as interaction

effects of time and temperature on flexural strength of sheet castings of this polymer.

Experiments with factors at two levels are useful for exploratory studies of the effects of many variables (or factors) on a response. Such studies are valuable in product development, for instance. Box, Hunter, and Hunter (1978) have extensive coverage of such experiments. For a general discussion of experiments with three or more factors, see a book on experimental design such as that by Kirk (1982).

In Chapter 14, we talk about comparing variances.

Summary of Chapter 13

Two-way, or two-factor, analysis of variance refers to parametric analysis of a two-factor, balanced, completely randomized, factorial experiment. We can test hypotheses about the effects of each of the factors on the response. With replication, we can also test for interaction effects of the two factors on the response variable.

There is an interaction effect of two variables on the response if the relative effects of levels of one variable depend on the level of the other variable. If there is an interaction effect of the two factors on response, then we say that each factor has an effect on response. (Note that interaction effect does *not* refer to any effect the two factors have on each other, but rather to their mutual effect on the response.)

Simple graphs make experimental results easy to interpret when each factor has just two levels. Such experiments are useful in exploratory studies that proceed in steps. The design of each successive experiment depends on the outcome of the previous experiments.

Minitab Appendix for Chapter 13

Finding Descriptive Statistics Within Levels of Two or More Variables

We can use the TABLE command with the STATS subcommand to print the mean and standard deviation of the responses within each combination of factors in a factorial experiment. Suppose our worksheet contains the data from Example 13-1. Niacin enrichment level is in column 1 (named NIALEVEL) and code for laboratory is in column 2 (named LAB). Niacin measurements are in column 3 (named MEASURE). If we use the command

```
MTB>   table 'nialevel' 'lab';
SUBC>  stats 'measure'.
```

we get the output in Figure M13-1.

```
ROWS: nialevel      COLUMNS: lab

              1         2         3         4         5         6       ALL

0             3         3         3         3         3         3        18
         3.5167    3.8333    3.9333    4.2300    4.0000    3.7467    3.8767
         0.1150    0.0577    0.2802    0.3236    0.3464    0.0611    0.2992

2             3         3         3         3         3         3        18
         5.2200    5.5667    5.6633    6.2533    5.8000    5.9033    5.7344
         0.1997    0.2517    0.1856    0.3573    0.1000    0.2628    0.3826

4             3         3         3         3         3         3        18
         6.9467    7.5667    7.1000    7.9267    7.3333    7.5900    7.4106
         0.4692    0.4509    0.6245    0.5178    0.3214    0.0854    0.5033

ALL           9         9         9         9         9         9        54
         5.2278    5.6556    5.5656    6.1367    5.7111    5.7467    5.6739
         1.5081    1.6387    1.4182    1.6417    1.4650    1.6743    1.5098


 CELL CONTENTS --
         measure:N
                 MEAN
                 STD DEV
```

FIGURE M13-1 Table of means and standard deviations of niacin measurements within combinations of NIALEVEL and LAB

Minitab prints the number of nonmissing values, the mean, and the standard deviation of MEASURE values within each combination of NIALEVEL and LAB. The variables in the main TABLE command must be classification variables, taking integer values between -9999 and $+9999$ or missing values. We can specify more than two classification variables in the main TABLE command if we have more than two factors to consider. Two or more variables can be listed in the STATS subcommand; the statistics will be printed for each variable listed.

Carrying Out a Two-Factor Analysis of Variance

The TWOWAY command performs some of the calculations for two-way analysis of variance. This command requires the same number of nonmissing observations within each combination of the two factors. The command

```
MTB> twoway 'measure' 'nialevel' 'lab' c4 c5
MTB> name c4 'resid' c5 'predict'
```

produces the two-way analysis of variance table in Figure M13-2. The residuals are stored in column 4, named RESID. The estimated or predicted values of the observations based on the two-way analysis of variance model are stored in column 5, named PREDICT.

We have to calculate the test statistics ourselves. In the following sequence of commands, we calculate and print the test statistic K1 and p-value K2 for the hypotheses about effects of the variable NIALEVEL on response, the test statistic K3 and p-value K4 for the hypotheses about effects of LAB on

```
ANALYSIS OF VARIANCE  measure

SOURCE         DF         SS         MS
nialevel        2    112.494     56.247
lab             5      3.887      0.777
INTERACTION    10      0.709      0.071
ERROR          36      3.720      0.103
TOTAL          53    120.811
```

FIGURE M13-2 Two-way analysis of variance table for Example 13-1

response, and the test statistic K5 and p-value K6 for the hypotheses about interaction effects of NIALEVEL and LAB on response:

```
MTB>   let k1=56.247/0.103
MTB>   cdf k1 k2;
SUBC>  f 2 36.
MTB>   let k2=1-k2
MTB>   let k3=0.777/0.103
MTB>   cdf k3 k4;
SUBC>  f 5 36.
MTB>   let k4=1-k4
MTB>   let k5=0.071/0.103
MTB>   cdf k5 k6;
SUBC>  f 10 36.
MTB>   let k6=1-k6
MTB>   print k1-k6
K1     546.087
K2     0
K3     7.54369
K4     0.000062466
K5     0.689320
K6     0.727334
```

These values for test statistics and p-values do not exactly equal those we found in Example 13-1, because we had to use the rounded values printed by Minitab. We can use the residuals stored in column 4 and the predicted values stored in column 5 in plots to check model assumptions. We might like to print the residuals and predicted values within each combination of our two factors. To do this, we use the TABLE command with the DATA subcommand:

```
MTB>   table 'nialevel' 'lab';
SUBC>  data 'resid' 'predict'.
```

Minitab prints all the values of each variable listed in the DATA subcommand, within each combination of levels of the variables listed in the main TABLE command. The results are shown in Figure M13-3.

We can use the TABLE command with the DATA subcommand to print the values of the response variable within combinations of levels of the factors, if we wish.

```
ROWS: nialevel      COLUMNS: lab

              1         2         3         4         5         6

0  -0.11667  -0.03333  -0.27333   0.14000   0.20000   0.01333
    0.11333  -0.03333  -0.01333  -0.37000  -0.40000  -0.06667
    0.00333   0.06667   0.28667   0.23000   0.20000   0.05333
     3.5167    3.8333    3.9333    4.2300    4.0000    3.7467
     3.5167    3.8333    3.9333    4.2300    4.0000    3.7467
     3.5167    3.8333    3.9333    4.2300    4.0000    3.7467

2  -0.22000  -0.26667   0.01667   0.27667   0.00000   0.15667
    0.05000   0.03333  -0.19333  -0.40333  -0.10000  -0.30333
    0.17000   0.23333   0.17667   0.12667   0.10000   0.14667
     5.2200    5.5667    5.6633    6.2533    5.8000    5.9033
     5.2200    5.5667    5.6633    6.2533    5.8000    5.9033
     5.2200    5.5667    5.6633    6.2533    5.8000    5.9033

4  -0.40667  -0.46667   0.20000   0.39333   0.36667   0.01000
    0.51333   0.03333  -0.70000  -0.58667  -0.23333  -0.09000
   -0.10667   0.43333   0.50000   0.19333  -0.13333   0.08000
     6.9467    7.5667    7.1000    7.9267    7.3333    7.5900
     6.9467    7.5667    7.1000    7.9267    7.3333    7.5900
     6.9467    7.5667    7.1000    7.9267    7.3333    7.5900

CELL CONTENTS --
          resid:DATA
        predict:DATA
```

FIGURE M13-3 Residuals and predicted values listed for each combination of NIALEVEL and LAB in Example 13-1

Exercises for Chapter 13

For each exercise, plot the observations in any ways that seem helpful. Describe the population(s) sampled, whether real or hypothetical. For each procedure, describe the assumptions that make the analysis appropriate. Do these assumptions seem reasonable? What additional information would you like to have about the experiment? Discuss the results of your analysis.

EXERCISE 13-1

A softball player wanted to see how far he could hit two brands of softball with two different bats. The two bats were each 34 inches long, and weighed 34 ounces. The balls were all new. A friend placed the balls in a pitching machine. The player did not know which brand of ball he was hitting. He did, of course, know the type of bat he was using for each trial. He hit four balls of each brand with each of the two bats. He and his friend measured the distance (in feet) he hit each ball. The results are shown below (Shaughnessy, 1988). The numbers in parentheses give the order in which the balls were hit.

	Wooden bat	Aluminum bat
Dudley Thunder	242 (9), 230 (11), 250 (12), 242 (8)	270 (4), 282 (5), 265 (6), 277 (3)
Worth Red Dot	258 (15), 264 (10), 265 (7), 275 (14)	290 (1), 318 (16), 302 (13), 310 (2)

a. Plot the observations. Include a plot similar to Figure 13-8.

b. From your plots, does there appear to be an interaction effect of type of ball and bat on distance hit?

c. Go through the steps for two-way analysis of variance.

d. Use residual plots to decide whether the assumptions for the analysis seer reasonable.

e. Plot residuals versus run order. Does there appear to be a time effect o distance hit?

f. Discuss your findings.

EXERCISE 13-2 The supervisor of a resin pilot plant wanted to study the effects of pH an temperature during formulation on the optical density of a polymer latex. Op tical density (or absorbance) is a characteristic of a polymer latex that manu facturers generally want to minimize. A skilled technician prepared sample of polymer latex with two levels of pH at each of two temperatures, tw samples per factor combination. The technician used standard techniques o emulsion polymerization, keeping all other conditions as constant as possible The supervisor used a random process to determine the order in which th samples were prepared. A different technician measured optical density fo each sample, using a standard electrophotometric instrument. This technicia was unaware of the experiment or the conditions under which the sample were prepared. The measured values of optical density (units not given) ar shown below (Gasper, 1988; with permission of ICI Resins US, a business uni of ICI Americas, Inc.). The number in parentheses is the order of the run.

	85 °C		95 °C	
pH 9.0	56.6 (6)	38.9 (7)	39.0 (4)	37.5 (5)
pH 9.3	63.0 (1)	96.8 (3)	33.0 (2)	33.3 (8)

a. Plot the observations in any ways that seem helpful. Include a plot similar to Figure 13-8.

b. From your plots, does there appear to be an interaction effect of tempera ture and pH on optical density?

c. Go through the steps for two-way analysis of variance.

d. Use residual plots to check assumptions. Include a plot of residuals versus run order. Do the assumptions of the analysis seem reasonable?

e. Find the reciprocal of each optical density observation. Plot these values as you did in part (a).

f. Go through the steps for two-way analysis of variance on the reciprocals of the optical density observations.

g. Use residual plots to check assumptions. Include a plot of residuals versus run order. Do the assumptions of the analysis seem reasonable?

h. Discuss your findings. If the supervisor wants to minimize optical density, what does this experiment suggest?

EXERCISE 13-3 In another part of the experiment discussed in Example 13-1, researchers divided samples of bran flakes into three groups. The bran flakes in the first group were not enriched with niacin. The bran flakes in the second group were enriched with 4 milligrams of niacin per 100 grams of flakes. The samples in the third group were enriched with 8 milligrams of niacin per 100 grams of flakes. The experimenters sent samples to the same six laboratories mentioned in Example 13-1. Laboratory workers divided a sample into three subsamples. They measured niacin in the sample on three different days, one subsample each day. The measurements of niacin (in milligrams per 100 grams) are shown below (Rice, 1988, pages 429–430; from Campbell and Pelletier, 1962).

Niacin enrichment	Laboratory					
	a	b	c	d	e	f
0	7.31	8.50	8.20	8.82	8.40	8.32
	7.85	8.50	8.25	8.76	8.60	8.25
	7.92	8.60	8.20	8.52	7.90	8.57
4	11.11	12.00	—	12.90	12.20	12.00
	11.00	13.10	11.68	12.00	11.60	12.40
	11.67	12.60	11.43	13.50	11.60	12.30
8	15.00	17.00	—	17.30	16.10	16.80
	17.00	17.50	16.20	17.60	16.10	16.60
	15.50	17.20	16.60	18.40	15.80	16.30

a. Plot the observations.

b. From the plots, does there appear to be an interaction effect of laboratory and niacin enrichment level on the niacin measurements?

c. Two observations are missing for laboratory c, one for each of two niacin enrichment levels. Therefore, this part of the experiment is not balanced; we do not have the same number of observations per factor combination. Analysis of an unbalanced experiment is more complicated than for a balanced experiment (and beyond the scope of our presentation). When only a few observations are missing, we can avoid a more complicated analysis. Replace a missing observation with the average of the other observations in the same factor combination. Subtract 1 from the residual degrees of freedom for each such substitution. Then proceed with the analysis as we have described. Use this procedure to replace each of the two missing values in this data set.

d. Go through the steps for two-way analysis of variance, accounting for the two missing values in the residual degrees of freedom.

e. Use residual plots to check the assumptions for the analysis.

f. Discuss your findings.

EXERCISE 13-4 This experiment is another phase of the experiment we discussed in Example 13-2. An engineer wanted to study the effect of two factors on strength of sheet

castings of a polymer. One factor was time in a polymerization bath with 2% catalyst. The other factor was temperature of the polymerization bath. He tried two times, 20 minutes and 60 minutes. He also tried two temperatures, 100° and 120° (temperature scale not given). The engineer treated samples of polymer under each of the four sets of conditions. Then he measured flexural strength (in pounds per square inch) on samples of sheet castings of the polymer. The results are shown below (Duncan, 1974, page 686; from Gore, 1947).

	100°					120°				
20 minutes	11,800	11,750	11,800	11,950	11,900	10,550	11,000	11,100	11,350	11,200
60 minutes	11,900	11,850	11,850	12,000	12,100	9,900	10,150	9,400	9,800	9,900

a. Plot the observations. Include a plot similar to Figure 13-8.

b. Does there appear to be an interaction effect of temperature and time in the bath on flexural strength?

c. Go through the steps for two-way analysis of variance.

d. Use residual plots to decide whether the assumptions for the analysis seem reasonable.

e. Discuss your findings.

EXERCISE 13-5 Researchers carried out an experiment to study retention of two forms of iron (Fe^{2+} and Fe^{3+}) given to mice at each of three concentrations (10.2, 1.2, and .3 millimolar). They randomly divided 108 mice into six groups, 18 mice per group. Each group of mice received one of the six combinations of type of iron and concentration. The researchers gave radioactively labeled iron to the mice orally. They used a counter to verify the amount of iron administered. They also used the counter to record the amount retained later (time not specified). Percentage of iron retained is shown below for each of the 108 mice (from an example in Rice, 1988, pages 356–357).

10.2 millimolar Fe^{3+}	.71	1.66	2.01	2.16	2.42	2.42	2.56	2.60	3.31
	3.64	3.74	3.74	4.39	4.50	5.07	5.26	8.15	8.24
1.2 millimolar Fe^{3+}	2.20	2.93	3.08	3.49	4.11	4.95	5.16	5.54	5.68
	6.25	7.25	7.90	8.85	11.96	15.54	15.89	18.30	18.59
.3 millimolar Fe^{3+}	2.25	3.93	5.08	5.82	5.84	6.89	8.50	8.56	9.44
	10.52	13.46	13.57	14.76	16.41	16.96	17.56	22.82	29.13
10.2 millimolar Fe^{2+}	2.20	2.69	3.54	3.75	3.83	4.08	4.27	4.53	5.32
	6.18	6.22	6.33	6.97	6.97	7.52	8.36	11.65	12.45
1.2 millimolar Fe^{2+}	4.04	4.16	4.42	4.93	5.49	5.77	5.86	6.28	6.97
	7.06	7.78	9.23	9.34	9.91	13.46	18.40	23.89	26.39
.3 millimolar Fe^{2+}	2.71	5.43	6.38	6.38	8.32	9.04	9.56	10.01	10.08
	10.62	13.80	15.99	17.90	18.25	19.32	19.87	21.60	22.25

a. Plot the observations.

b. From your plots, does there appear to be an interaction effect of form of iron and concentration on percentage retained?

c. Go through the steps for two-way analysis of variance. Use residual plots to check the assumptions for the analysis. Do the assumptions seem reasonable?

d. Plot the logarithm of percentage of iron retained.

e. Go through the steps for two-way analysis of variance, using the logarithms of the observations. Use the residual plots to check the assumptions of the analysis. Do these assumptions seem reasonable?

f. Compare your answers to parts (c) and (e). Discuss your findings.

EXERCISE 13-6 An engineering student carried out a study of the effects of bathing on fecal and total coliform (colon bacillus) bacteria in the bath water. Eight males and eight females participated. Each volunteer took a bath in a 100-gallon polyethylene tub, with dechlorinated 38 °C tap water. Two factors studied were time since the volunteer's last bath and the level of his or her activity in the bath. After a volunteer had been in the tub for 15 minutes, the engineering student recorded two response variables: change in fecal coliform concentration and change in total coliform concentration (organisms per 100 milliliters). The results are shown below (Box, Hunter, and Hunter, 1978, pages 435–436; from Drew, 1971).

Activity	Time since last bath	Females		Males	
Change in fecal coliform concentration after 15 minutes (organisms per 100 ml)					
Lethargic	1 hour	1	2	153	96
Lethargic	24 hours	12	37	129	390
Vigorous	1 hour	16	21	143	300
Vigorous	24 hours	4	2	113	280
Change in total coliform concentration after 15 minutes (organisms per 100 ml)					
Lethargic	1 hour	3	10	426	147
Lethargic	24 hours	57	280	250	1,470
Vigorous	1 hour	323	33	580	665
Vigorous	24 hours	183	10	650	675

a. Plot the values of change in fecal coliform concentration for females. Does there appear to be an interaction effect of time since last bath and bathing activity on change in fecal coliform concentration?

b. Repeat part (a) after taking the logarithm of each observation. Do the assumptions for two-way analysis of variance seem better met by the transformed observations?

c. Analyze the logarithm of change in fecal coliform concentration for females. Discuss your findings.

d. Repeat parts (a) and (b) for males.

e. Repeat part (c) for males.

f. Compare your findings in parts (a), (b), and (c) for the female volunteers with your findings in parts (d) and (e) for the male volunteers.

g. Plot the values of change in total coliform concentration for females. Does there appear to be an interaction effect of time since last bath and bathing activity on the observations?

h. Repeat part (g) after taking the logarithm of each observation. Do the assumptions for two-way analysis of variance seem better met by the transformed observations?

i. Analyze the logarithm of change in total coliform concentration for females. Discuss your findings.

j. Repeat parts (g) and (h) for males.

k. Repeat part (i) for males.

l. Compare your findings in parts (g), (h), and (i) for the female volunteers with your findings in parts (j) and (k) for the males.

EXERCISE 13-7 Researchers wanted to study four treatments in counteracting the effects of three poisons. They randomly assigned 48 animals to treatment/poison combinations, four animals per combination. The researchers measured survival time in hours for each animal. The results are shown below (Rice, 1988, pages 432–433; Box, Hunter, and Hunter, 1978, page 228; from Box and Cox, 1964).

	Poison 1				Poison 2				Poison 3			
Treatment a	3.1	4.5	4.6	4.3	3.6	2.9	4.0	2.3	2.2	2.1	1.8	2.3
Treatment b	8.2	11.0	8.8	7.2	9.2	6.1	4.9	12.4	3.0	3.7	3.8	2.9
Treatment c	4.3	4.5	6.3	7.6	4.4	3.5	3.1	4.0	2.3	2.5	2.4	2.2
Treatment d	4.5	7.1	6.6	6.2	5.6	10.0	7.1	3.8	3.0	3.6	3.1	3.3

a. Plot the observations.

b. From your plots, does there appear to be an interaction effect of poison and treatment on survival?

c. Go through the steps for two-way analysis of variance. Use residual plots to check model assumptions. Do these assumptions seem reasonable?

d. You can interpret the reciprocal of a survival time (time until death) as a death rate (number of deaths per unit time). Take the reciprocal of each of the 48 survival times. Plot these transformed observations.

e. Go through the steps for two-way analysis of variance on the reciprocals of the observations. Use residual plots to check the assumptions for the analysis. Do these assumptions seem reasonable?

f. Compare your answers to parts (c) and (e). Discuss your findings.

EXERCISE 13-8 Managers were interested in the productivity of four technicians on each of five machines. They recorded the number of units produced by a technician on each machine, two different days. The results are shown below. Managers numbered the 84 working days required for the experiment, from 1 to 84. The day the observation was made is shown in parentheses next to the observation (from an exercise on page 285 in *Statistics for Experimenters* by G. E. P. Box, W. G. Hunter, and J. S. Hunter, 1978, John Wiley and Sons, Inc., New York).

	Machine 1	Machine 2	Machine 3	Machine 4	Machine 5
Technician a	18 (9), 17 (76)	17 (1), 13 (71)	16 (3), 17 (77)	15 (2), 17 (72)	17 (17), 18 (84)
Technician b	16 (11), 18 (77)	18 (3), 18 (73)	17 (7), 19 (70)	21 (4), 22 (74)	16 (10), 18 (72)
Technician c	17 (22), 20 (72)	20 (57), 16 (70)	20 (25), 16 (73)	16 (5), 16 (71)	14 (39), 13 (74)
Technician d	27 (3), 27 (73)	28 (2), 23 (78)	31 (33), 30 (72)	31 (6), 24 (75)	28 (7), 22 (82)

a. Plot the observations.

b. Does there appear to be an interaction effect of technician and machine on number of units produced?

c. Go through the steps for two-way analysis of variance. Use residual plots to check model assumptions. Do these assumptions seem reasonable?

d. Plot residuals versus time of observation. Is there a trend? If there were a trend, what would it suggest?

e. Discuss your findings.