13 Design and Analysis of Single-Factor Experiments: The Analysis of Variance

CHAPTER OUTLINE

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Design and conduct engineering experiments involving a single factor with an arbitrary number of levels
2. Understand how the analysis of variance is used to analyze the data from these experiments
3. Assess model adequacy with residual plots
4. Use multiple comparison procedures to identify specific differences between means
5. Make decisions about sample size in single-factor experiments
6. Understand the difference between fixed and random factors
7. Estimate variance components in an experiment involving random factors

8. **Understand the blocking principle and how it is used to isolate the effect of nuisance factors**
9. **Design and conduct experiments involving the randomized complete block design**

CD MATERIAL

10. **Use operating characteristic curves to make sample size decisions in single-factor random effects experiment**
11. **Use Tukey's test, orthogonal contrasts and graphical methods to identify specific differences between means.**

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-Text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also available for some of the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

13-1 DESIGNING ENGINEERING EXPERIMENTS

Experiments are a natural part of the engineering and scientific decision-making process. Suppose, for example, that a civil engineer is investigating the effects of different curing methods on the mean compressive strength of concrete. The experiment would consist of making up several test specimens of concrete using each of the proposed curing methods and then testing the compressive strength of each specimen. The data from this experiment could be used to determine which curing method should be used to provide maximum mean compressive strength.

If there are only two curing methods of interest, this experiment could be designed and analyzed using the statistical hypothesis methods for two samples introduced in Chapter 10. That is, the experimenter has a single **factor** of interest—curing methods—and there are only two **levels** of the factor. If the experimenter is interested in determining which curing method produces the maximum compressive strength, the number of specimens to test can be determined from the operating characteristic curves in Appendix Chart VI, and the *t*-test can be used to decide if the two means differ.

Many single-factor experiments require that more than two levels of the factor be considered. For example, the civil engineer may want to investigate five different curing methods. In this chapter we show how the **analysis of variance** (frequently abbreviated ANOVA) can be used for comparing means when there are more than two levels of a single factor. We will also discuss **randomization** of the experimental runs and the important role this concept plays in the overall experimentation strategy. In the next chapter, we will show how to design and analyze experiments with several factors.

Statistically based experimental design techniques are particularly useful in the engineering world for improving the performance of a manufacturing process. They also have extensive application in the development of new processes. Most processes can be described in terms of several **controllable variables,** such as temperature, pressure, and feed rate. By using designed experiments, engineers can determine which subset of the process variables has the greatest influence on process performance. The results of such an experiment can lead to

1. Improved process yield
2. Reduced variability in the process and closer conformance to nominal or target requirements
3. Reduced design and development time
4. Reduced cost of operation

Experimental design methods are also useful in **engineering design** activities, where new products are developed and existing ones are improved. Some typical applications of statistically designed experiments in engineering design include

1. Evaluation and comparison of basic design configurations
2. Evaluation of different materials
3. Selection of design parameters so that the product will work well under a wide variety of field conditions (or so that the design will be robust)
4. Determination of key product design parameters that affect product performance

The use of experimental design in the engineering design process can result in products that are easier to manufacture, products that have better field performance and reliability than their competitors, and products that can be designed, developed, and produced in less time.

Designed experiments are usually employed **sequentially.** That is, the first experiment with a complex system (perhaps a manufacturing process) that has many controllable variables is often a **screening experiment** designed to determine which variables are most important. Subsequent experiments are used to refine this information and determine which adjustments to these critical variables are required to improve the process. Finally, the objective of the experimenter is optimization, that is, to determine which levels of the critical variables result in the best process performance.

Every experiment involves a sequence of activities:

1. **Conjecture**—the original hypothesis that motivates the experiment.
2. **Experiment**—the test performed to investigate the conjecture.
3. **Analysis**—the statistical analysis of the data from the experiment.
4. **Conclusion**—what has been learned about the original conjecture from the experiment. Often the experiment will lead to a revised conjecture, and a new experiment, and so forth.

The statistical methods introduced in this chapter and Chapter 14 are essential to good experimentation. **All experiments are designed experiments;** unfortunately, some of them are poorly designed, and as a result, valuable resources are used ineffectively. Statistically designed experiments permit efficiency and economy in the experimental process, and the use of statistical methods in examining the data results in **scientific objectivity** when drawing conclusions.

13-2 THE COMPLETELY RANDOMIZED SINGLE-FACTOR EXPERIMENT

13-2.1 An Example

A manufacturer of paper used for making grocery bags is interested in improving the tensile strength of the product. Product engineering thinks that tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentrations of practical interest is between 5 and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5%, 10%, 15%, and 20%. They decide to make up six test specimens at each concentration level, using a pilot plant. All 24 specimens are tested on a laboratory tensile tester, in random order. The data from this experiment are shown in Table 13-1.

Table 13-1 Tensile Strength of Paper (psi)

Hardwood Concentration (%)	Observations						Totals	Averages
	1	2	3	4	5	6		
5	7	8	15	11	9	10	60	10.00
10	12	17	13	18	19	15	94	15.67
15	14	18	19	17	16	18	102	17.00
20	19	25	22	23	18	20	127	21.17
							383	15.96

This is an example of a completely randomized single-factor experiment with four levels of the factor. The levels of the factor are sometimes called **treatments,** and each treatment has six observations or **replicates.** The role of **randomization** in this experiment is extremely important. By randomizing the order of the 24 runs, the effect of any nuisance variable that may influence the observed tensile strength is approximately balanced out. For example, suppose that there is a warm-up effect on the tensile testing machine; that is, the longer the machine is on, the greater the observed tensile strength. If all 24 runs are made in order of increasing hardwood concentration (that is, all six 5% concentration specimens are tested first, followed by all six 10% concentration specimens, etc.), any observed differences in tensile strength could also be due to the warm-up effect.

It is important to graphically analyze the data from a designed experiment. Figure 13-1(a) presents box plots of tensile strength at the four hardwood concentration levels. This figure indicates that changing the hardwood concentration has an effect on tensile strength; specifically, higher hardwood concentrations produce higher observed tensile strength. Furthermore, the distribution of tensile strength at a particular hardwood level is reasonably symmetric, and the variability in tensile strength does not change dramatically as the hardwood concentration changes.

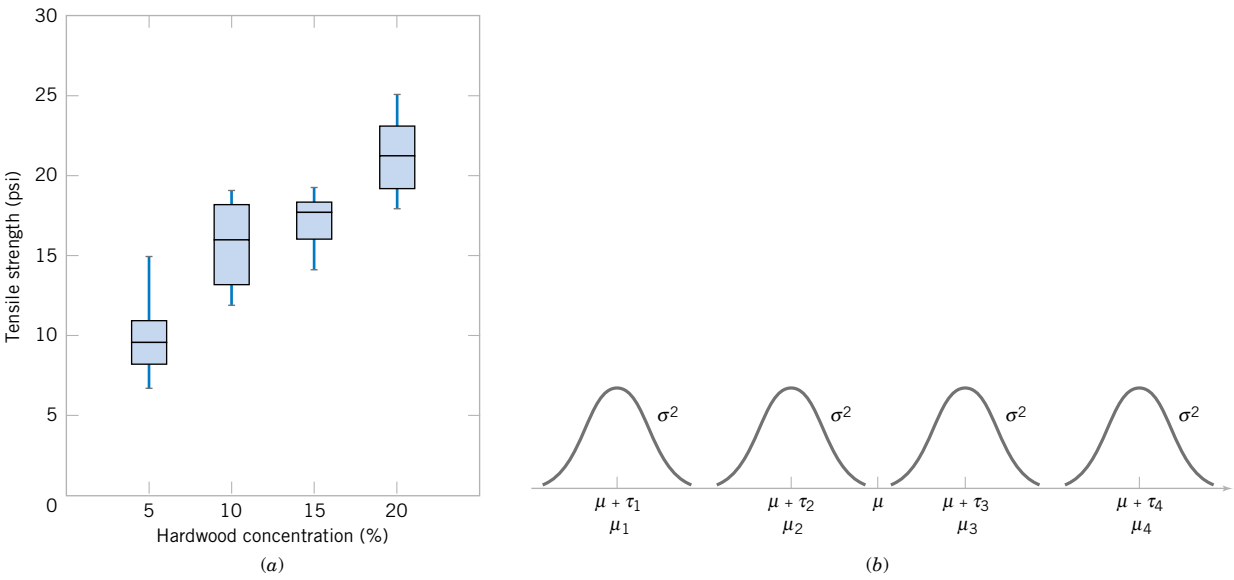


Figure 13-1 (a) Box plots of hardwood concentration data. (b) Display of the model in Equation 13-1 for the completely randomized single-factor experiment.

Graphical interpretation of the data is always useful. Box plots show the variability of the observations *within* a treatment (factor level) and the variability *between* treatments. We now discuss how the data from a single-factor randomized experiment can be analyzed statistically.

13-2.2 The Analysis of Variance

Suppose we have a different levels of a single factor that we wish to compare. Sometimes, each factor level is called a **treatment,** a very general term that can be traced to the early applications of experimental design methodology in the agricultural sciences. The **response** for each of the a treatments is a random variable. The observed data would appear as shown in Table 13-2. An entry in Table 13-2, say y_{ij}, represents the jth observation taken under treatment i. We initially consider the case in which there are an equal number of observations, n, on each treatment.

We may describe the observations in Table 13-2 by the **linear statistical model**

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases} \tag{13-1}$$

where Y_{ij} is a random variable denoting the (ij)th observation, μ is a parameter common to all treatments called the **overall mean,** τ_i is a parameter associated with the ith treatment called the ith **treatment effect,** and ϵ_{ij} is a random error component. Notice that the model could have been written as

$$Y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

where $\mu_i = \mu + \tau_i$ is the mean of the ith treatment. In this form of the model, we see that each treatment defines a population that has mean μ_i, consisting of the overall mean μ plus an effect τ_i that is due to that particular treatment. We will assume that the errors ϵ_{ij} are normally and independently distributed with mean zero and variance σ^2. Therefore, each treatment can be thought of as a normal population with mean μ_i and variance σ^2. See Fig. 13-1(b).

Equation 13-1 is the underlying model for a single-factor experiment. Furthermore, since we require that the observations are taken in random order and that the environment (often called the experimental units) in which the treatments are used is as uniform as possible, this experimental design is called a **completely randomized design.**

Table 13-2 Typical Data for a Single-Factor Experiment

Treatment	Observations				Totals	Averages
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots\vdots\vdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
					$y_{\cdot\cdot}$	$\bar{y}_{\cdot\cdot}$

The a factor levels in the experiment could have been chosen in two different ways. First, the experimenter could have specifically chosen the a treatments. In this situation, we wish to test hypotheses about the treatment means, and conclusions cannot be extended to similar treatments that were not considered. In addition, we may wish to estimate the treatment effects. This is called the **fixed-effects model.** Alternatively, the a treatments could be a random sample from a larger population of treatments. In this situation, we would like to be able to extend the conclusions (which are based on the sample of treatments) to all treatments in the population, whether or not they were explicitly considered in the experiment. Here the treatment effects τ_i are random variables, and knowledge about the particular ones investigated is relatively unimportant. Instead, we test hypotheses about the variability of the τ_i and try to estimate this variability. This is called the **random effects,** or **components of variance,** model.

In this section we develop the **analysis of variance** for the fixed-effects model. The analysis of variance is not new to us; it was used previously in the presentation of regression analysis. However, in this section we show how it can be used to test for equality of treatment effects. In the fixed-effects model, the treatment effects τ_i are usually defined as deviations from the overall mean μ, so that

$$\sum_{i=1}^{a} \tau_i = 0 \tag{13-2}$$

Let $y_{i\cdot}$ represent the total of the observations under the ith treatment and $\bar{y}_{i\cdot}$ represent the average of the observations under the ith treatment. Similarly, let $y_{\cdot\cdot}$ represent the grand total of all observations and $\bar{y}_{\cdot\cdot}$ represent the grand mean of all observations. Expressed mathematically,

$$\begin{aligned} y_{i\cdot} &= \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{i\cdot} = y_{i\cdot}/n \qquad i = 1, 2, \ldots, a \\ y_{\cdot\cdot} &= \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/N \end{aligned} \tag{13-3}$$

where $N = an$ is the total number of observations. Thus, the "dot" subscript notation implies summation over the subscript that it replaces.

We are interested in testing the equality of the a treatment means $\mu_1, \mu_2, \ldots, \mu_a$. Using Equation 13-2, we find that this is equivalent to testing the hypotheses

$$\begin{aligned} &H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0 \\ &H_1: \tau_i \neq 0 \quad \text{for at least one } i \end{aligned} \tag{13-4}$$

Thus, if the null hypothesis is true, each observation consists of the overall mean μ plus a realization of the random error component ϵ_{ij}. This is equivalent to saying that all N observations are taken from a normal distribution with mean μ and variance σ^2. Therefore, if the null hypothesis is true, changing the levels of the factor has no effect on the mean response.

The ANOVA partitions the total variability in the sample data into two component parts. Then, the test of the hypothesis in Equation 13-4 is based on a comparison of two independent estimates of the population variance. The total variability in the data is described by the **total sum of squares**

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2$$

The partition of the total sum of squares is given in the following definition.

Definition

The **sum of squares identity** is

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 \qquad (13\text{-}5)$$

or symbolically

$$SS_T = SS_{\text{Treatments}} + SS_E \qquad (13\text{-}6)$$

The identity in Equation 13-5 (which is developed in Section 13-4.4 on the CD) shows that the total variability in the data, measured by the total corrected sum of squares SS_T, can be partitioned into a sum of squares of differences between treatment means and the grand mean denoted $SS_{\text{Treatments}}$ and a sum of squares of differences of observations within a treatment from the treatment mean denoted SS_E. Differences between observed treatment means and the grand mean measure the differences between treatments, while differences of observations within a treatment from the treatment mean can be due only to random error.

We can gain considerable insight into how the analysis of variance works by examining the expected values of $SS_{\text{Treatments}}$ and SS_E. This will lead us to an appropriate statistic for testing the hypothesis of no differences among treatment means (or all $\tau_i = 0$).

The expected value of the treatment sum of squares is

$$E(SS_{\text{Treatments}}) = (a-1)\sigma^2 + n\sum_{i=1}^{a}\tau_i^2$$

and the expected value of the error sum of squares is

$$E(SS_E) = a(n-1)\sigma^2$$

There is also a partition of the number of degrees of freedom that corresponds to the sum of squares identity in Equation 13-5. That is, there are $an = N$ observations; thus, SS_T has $an - 1$ degrees of freedom. There are a levels of the factor, so $SS_{\text{Treatments}}$ has $a - 1$ degrees of freedom. Finally, within any treatment there are n replicates providing $n - 1$ degrees of freedom with which to estimate the experimental error. Since there are a treatments, we have $a(n-1)$ degrees of freedom for error. Therefore, the degrees of freedom partition is

$$an - 1 = a - 1 + a(n-1)$$

The ratio

$$MS_{\text{Treatments}} = SS_{\text{Treatments}}/(a-1)$$

is called the **mean square for treatments.** Now if the null hypothesis $H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$ is true, $MS_{\text{Treatments}}$ is an unbiased estimator of σ^2 because $\sum_{i=1}^{a}\tau_i = 0$. However, if H_1 is true, $MS_{\text{Treatments}}$ estimates σ^2 plus a positive term that incorporates variation due to the systematic difference in treatment means.

Note that the **error mean square**

$$MS_E = SS_E/[a(n-1)]$$

is an unbiased estimator of σ^2 regardless of whether or not H_0 is true. We can also show that $MS_{\text{Treatments}}$ and MS_E are independent. Consequently, we can show that if the null hypothesis H_0 is true, the ratio

$$F_0 = \frac{SS_{\text{Treatments}}/(a-1)}{SS_E/[a(n-1)]} = \frac{MS_{\text{Treatments}}}{MS_E} \tag{13-7}$$

has an F-distribution with $a-1$ and $a(n-1)$ degrees of freedom. Furthermore, from the expected mean squares, we know that MS_E is an unbiased estimator of σ^2. Also, under the null hypothesis, $MS_{\text{Treatments}}$ is an unbiased estimator of σ^2. However, if the null hypothesis is false, the expected value of $MS_{\text{Treatments}}$ is greater than σ^2. Therefore, under the alternative hypothesis, the expected value of the numerator of the test statistic (Equation 13-7) is greater than the expected value of the denominator. Consequently, we should reject H_0 if the statistic is large. This implies an upper-tail, one-tail critical region. Therefore, we would reject H_0 if $f_0 > f_{\alpha,a-1,a(n-1)}$ where f_0 is the computed value of F_0 from Equation 13-7.

Efficient computational formulas for the sums of squares may be obtained by expanding and simplifying the definitions of $SS_{\text{Treatments}}$ and SS_T. This yields the following results.

Definition

The sums of squares computing formulas for the ANOVA with equal sample sizes in each treatment are

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{13-8}$$

and

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_i^2}{n} - \frac{y_{..}^2}{N} \tag{13-9}$$

The error sum of squares is obtained by subtraction as

$$SS_E = SS_T - SS_{\text{Treatments}} \tag{13-10}$$

The computations for this test procedure are usually summarized in tabular form as shown in Table 13-3. This is called an **analysis of variance** (or **ANOVA**) **table.**

Table 13-3 The Analysis of Variance for a Single-Factor Experiment, Fixed-Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a-1$	$MS_{\text{Treatments}}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Error	SS_E	$a(n-1)$	MS_E	
Total	SS_T	$an-1$		

EXAMPLE 13-1

Consider the paper tensile strength experiment described in Section 13-2.1. We can use the analysis of variance to test the hypothesis that different hardwood concentrations do not affect the mean tensile strength of the paper.

The hypotheses are

$$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$
$$H_1: \tau_i \neq 0 \text{ for at least one } i \quad .$$

We will use $\alpha = 0.01$. The sums of squares for the analysis of variance are computed from Equations 13-8, 13-9, and 13-10 as follows:

$$\begin{aligned} SS_T &= \sum_{i=1}^{4}\sum_{j=1}^{6} y_{ij}^2 - \frac{y_{..}^2}{N} \\ &= (7)^2 + (8)^2 + \cdots + (20)^2 - \frac{(383)^2}{24} = 512.96 \\ SS_{\text{Treatments}} &= \sum_{i=1}^{4} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N} \\ &= \frac{(60)^2 + (94)^2 + (102)^2 + (127)^2}{6} - \frac{(383)^2}{24} = 382.79 \\ SS_E &= SS_T - SS_{\text{Treatments}} \\ &= 512.96 - 382.79 = 130.17 \end{aligned}$$

The ANOVA is summarized in Table 13-4. Since $f_{0.01,3,20} = 4.94$, we reject H_0 and conclude that hardwood concentration in the pulp significantly affects the mean strength of the paper. We can also find a P-value for this test statistic as follows:

$$P = P(F_{3,20} > 19.60) \simeq 3.59 \times 10^{-6}$$

Since $P \simeq 3.59 \times 10^{-6}$ is considerably smaller than $\alpha = 0.01$, we have strong evidence to conclude that H_0 is not true.

Minitab Output

Many software packages have the capability to analyze data from designed experiments using the analysis of variance. Table 13-5 presents the output from the Minitab one-way analysis of variance routine for the paper tensile strength experiment in Example 13-1. The results agree closely with the manual calculations reported previously in Table 13-4.

Table 13-4 ANOVA for the Tensile Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Hardwood concentration	382.79	3	127.60	19.60	3.59 E-6
Error	130.17	20	6.51		
Total	512.96	23			

Table 13-5 Minitab Analysis of Variance Output for Example 13-1

One-Way ANOVA: Strength versus CONC

Analysis of Variance for Strength

Source	DF	SS	MS	F	P
Conc	3	382.79	127.60	19.61	0.000
Error	20	130.17	6.51		
Total	23	512.96			

Level	N	Mean	StDev
5	6	10.000	2.828
10	6	15.667	2.805
15	6	17.000	1.789
20	6	21.167	2.639

Pooled StDev = 2.551

Individual 95% CIs For Mean Based on Pooled StDev

```
----+---------+---------+---------+--
(---*---)
          (---*---)
               (---*---)
                    (---*---)
----+---------+---------+---------+--
   10.0      15.0      20.0      25.0
```

Fisher's pairwise comparisons

Family error rate = 0.192
Individual error rate = 0.0500

Critical value = 2.086

Intervals for (column level mean) − (row level mean)

	5	10	15
10	−8.739 −2.594		
15	−10.072 −3.928	−4.406 1.739	
20	−14.239 −8.094	−8.572 −2.428	−7.239 −1.094

The Minitab output also presents 95% **confidence intervals** on each individual treatment mean. The mean of the ith treatment is defined as

$$\mu_i = \mu + \tau_i \quad i = 1, 2, \ldots, a$$

A point estimator of μ_i is $\hat{\mu}_i = \bar{Y}_{i.}$. Now, if we assume that the errors are normally distributed, each treatment average is normally distributed with mean μ_i and variance σ^2/n. Thus, if σ^2 were known, we could use the normal distribution to construct a CI. Using MS_E as an estimator of σ^2 (The square root of MS_E is the "Pooled StDev" referred to in the Minitab output), we would base the CI on the t-distribution, since

$$T = \frac{\bar{Y}_{i.} - \mu_i}{\sqrt{MS_E/n}}$$

has a t-distribution with $a(n - 1)$ degrees of freedom. This leads to the following definition of the confidence interval.

Definition

A $100(1-\alpha)$ percent confidence interval on the mean of the ith treatment μ_i is

$$\bar{y}_{i\cdot} - t_{\alpha/2,a(n-1)}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_{i\cdot} + t_{\alpha/2,a(n-1)}\sqrt{\frac{MS_E}{n}} \tag{13-11}$$

Equation 13-11 is used to calculate the 95% CIs shown graphically in the Minitab output of Table 13-5. For example, at 20% hardwood the point estimate of the mean is $\bar{y}_{4\cdot} = 21.167$, $MS_E = 6.51$, and $t_{0.025,20} = 2.086$, so the 95% CI is

$$[\bar{y}_{4\cdot} \pm t_{0.025,20}\sqrt{MS_E/n}]$$
$$[21.167 \pm (2.086)\sqrt{6.51/6}]$$

or

$$19.00 \text{ psi} \le \mu_4 \le 23.34 \text{ psi}$$

It can also be interesting to find confidence intervals on the difference in two treatment means, say, $\mu_i - \mu_j$. The point estimator of $\mu_i - \mu_j$ is $\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}$, and the variance of this estimator is

$$V(\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{2\sigma^2}{n}$$

Now if we use MS_E to estimate σ^2,

$$T = \frac{\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} - (\mu_i - \mu_j)}{\sqrt{2MS_E/n}}$$

has a t-distribution with $a(n-1)$ degrees of freedom. Therefore, a CI on $\mu_i - \mu_j$ may be based on the t-distribution.

Definition

A $100(1-\alpha)$ percent confidence interval on the difference in two treatment means $\mu_i - \mu_j$ is

$$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} - t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}} \le \mu_i - \mu_j \le \bar{y}_{i\cdot} - \bar{y}_{j\cdot} + t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}} \tag{13-12}$$

A 95% CI on the difference in means $\mu_3 - \mu_2$ is computed from Equation 13-12 as follows:

$$[\bar{y}_{3\cdot} - \bar{y}_{2\cdot} \pm t_{0.025,20}\sqrt{2MS_E/n}]$$
$$[17.00 - 15.67 \pm (2.086)\sqrt{2(6.51)/6}]$$

or

$$-1.74 \le \mu_3 - \mu_2 \le 4.40$$

Since the CI includes zero, we would conclude that there is no difference in mean tensile strength at these two particular hardwood levels.

The bottom portion of the computer output in Table 13-5 provides additional information concerning which specific means are different. We will discuss this in more detail in Section 13-2.3.

An Unbalanced Experiment

In some single-factor experiments, the number of observations taken under each treatment may be different. We then say that the design is **unbalanced.** In this situation, slight modifications must be made in the sums of squares formulas. Let n_i observations be taken under treatment i ($i = 1, 2, \ldots, a$), and let the total number of observations $N = \sum_{i=1}^{a} n_i$. The computational formulas for SS_T and $SS_{\text{Treatments}}$ are as shown in the following definition.

Definition

The sums of squares computing formulas for the ANOVA with unequal sample sizes n_i in each treatment are

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{13-13}$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} \tag{13-14}$$

and

$$SS_E = SS_T - SS_{\text{Treatments}} \tag{13-15}$$

Choosing a balanced design has two important advantages. First, the ANOVA is relatively insensitive to small departures from the assumption of equality of variances if the sample sizes are equal. This is not the case for unequal sample sizes. Second, the power of the test is maximized if the samples are of equal size.

13-2.3 Multiple Comparisons Following the ANOVA

When the null hypothesis H_0: $\tau_1 = \tau_2 = \cdots = \tau_a = 0$ is rejected in the ANOVA, we know that some of the treatment or factor level means are different. However, the ANOVA doesn't identify which means are different. Methods for investigating this issue are called **multiple comparisons methods.** Many of these procedures are available. Here we describe a very simple one, **Fisher's least significant difference (LSD) method.** In Section 13-2.4 on the CD, we describe three other procedures. Montgomery (2001) presents these and other methods and provides a comparative discussion.

The Fisher LSD method compares all pairs of means with the null hypotheses H_0: $\mu_i = \mu_j$ (for all $i \neq j$) using the t-statistic

$$t_0 = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{\frac{2MS_E}{n}}}$$

Assuming a two-sided alternative hypothesis, the pair of means μ_i and μ_j would be declared significantly different if

$$|\bar{y}_{i.} - \bar{y}_{j.}| > \text{LSD}$$

where LSD, the **least significant difference,** is

$$\text{LSD} = t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}} \tag{13-16}$$

If the sample sizes are different in each treatment, the LSD is defined as

$$\text{LSD} = t_{\alpha/2,N-a}\sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

EXAMPLE 13-2

We will apply the Fisher LSD method to the hardwood concentration experiment. There are $a = 4$ means, $n = 6$, $MS_E = 6.51$, and $t_{0.025,20} = 2.086$. The treatment means are

$$\begin{aligned}\bar{y}_{1.} &= 10.00 \text{ psi}\\ \bar{y}_{2.} &= 15.67 \text{ psi}\\ \bar{y}_{3.} &= 17.00 \text{ psi}\\ \bar{y}_{4.} &= 21.17 \text{ psi}\end{aligned}$$

The value of LSD is $\text{LSD} = t_{0.025,20}\sqrt{2MS_E/n} = 2.086\sqrt{2(6.51)/6} = 3.07$. Therefore, any pair of treatment averages that differs by more than 3.07 implies that the corresponding pair of treatment means are different.

The comparisons among the observed treatment averages are as follows:

$$\begin{aligned}4 \text{ vs. } 1 &= 21.17 - 10.00 = 11.17 > 3.07\\ 4 \text{ vs. } 2 &= 21.17 - 15.67 = 5.50 > 3.07\\ 4 \text{ vs. } 3 &= 21.17 - 17.00 = 4.17 > 3.07\\ 3 \text{ vs. } 1 &= 17.00 - 10.00 = 7.00 > 3.07\\ 3 \text{ vs. } 2 &= 17.00 - 15.67 = 1.33 < 3.07\\ 2 \text{ vs. } 1 &= 15.67 - 10.00 = 5.67 > 3.07\end{aligned}$$

From this analysis, we see that there are significant differences between all pairs of means except 2 and 3. This implies that 10 and 15% hardwood concentration produce approximately the same tensile strength and that all other concentration levels tested produce different tensile strengths. It is often helpful to draw a graph of the treatment means, such as in Fig. 13-2, with the means that are *not* different underlined. This graph clearly reveals the results of the experiment and shows that 20% hardwood produces the maximum tensile strength.

The Minitab output in Table 13-5 shows the Fisher LSD method under the heading "Fisher's pairwise comparisons." The critical value reported is actually the value of $t_{0.025,20} =$

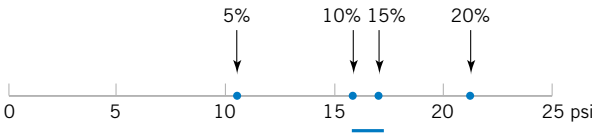


Figure 13-2 Results of Fisher's LSD method in Example 13-2.

2.086. Minitab implements Fisher's LSD method by computing **confidence intervals** on all pairs of treatment means using Equation 13-12. The lower and upper 95% confidence limits are shown at the bottom of the table. Notice that the only pair of means for which the confidence interval includes zero is for μ_{10} and μ_{15}. This implies that μ_{10} and μ_{15} are not significantly different, the same result found in Example 13-2.

Table 13-5 also provides a "family error rate," equal to 0.192 in this example. When all possible pairs of means are tested, the probability of at least one type I error can be much greater than for a single test. We can interpret the family error rate as follows. The probability is $1 - 0.192 = 0.808$ that there are no type I errors in the six comparisons. The family error rate in Table 13-5 is based on the distribution of the range of the sample means. See Montgomery (2001) for details. Alternatively, Minitab permits you to specify a family error rate and will then calculate an individual error rate for each comparison.

13-2.4 More About Multiple Comparisons (CD Only)

13-2.5 Residual Analysis and Model Checking

The analysis of variance assumes that the observations are normally and independently distributed with the same variance for each treatment or factor level. These assumptions should be checked by examining the residuals. A **residual** is the difference between an observation y_{ij} and its estimated (or fitted) value from the statistical model being studied, denoted as $\hat{y}_{ij}$. For the completely randomized design $\hat{y}_{ij} = \bar{y}_{i\cdot}$ and each residual is $e_{ij} = y_{ij} - \bar{y}_{i\cdot}$, that is, the difference between an observation and the corresponding observed treatment mean. The residuals for the paper tensile strength experiment are shown in Table 13-6. Using $\bar{y}_{i\cdot}$ to calculate each residual essentially removes the effect of hardwood concentration from the data; consequently, the residuals contain information about unexplained variability.

The normality assumption can be checked by constructing a **normal probability plot** of the residuals. To check the assumption of equal variances at each factor level, plot the residuals against the factor levels and compare the spread in the residuals. It is also useful to plot the residuals against $\bar{y}_{i\cdot}$ (sometimes called the fitted value); the variability in the residuals should not depend in any way on the value of $\bar{y}_{i\cdot}$. Most statistics software packages will construct these plots on request. When a pattern appears in these plots, it usually suggests the need for a transformation, that is, analyzing the data in a different metric. For example, if the variability in the residuals increases with $\bar{y}_{i\cdot}$, a transformation such as $\log y$ or $\sqrt{y}$ should be considered. In some problems, the dependency of residual scatter on the observed mean $\bar{y}_{i\cdot}$ is very important information. It may be desirable to select the factor level that results in maximum response; however, this level may also cause more variation in response from run to run.

The independence assumption can be checked by plotting the residuals against the time or run order in which the experiment was performed. A pattern in this plot, such as sequences of positive and negative residuals, may indicate that the observations are not independent.

Table 13-6 Residuals for the Tensile Strength Experiment

Hardwood Concentration (%)	Residuals					
5	−3.00	−2.00	5.00	1.00	−1.00	0.00
10	−3.67	1.33	−2.67	2.33	3.33	−0.67
15	−3.00	1.00	2.00	0.00	−1.00	1.00
20	−2.17	3.83	0.83	1.83	−3.17	−1.17

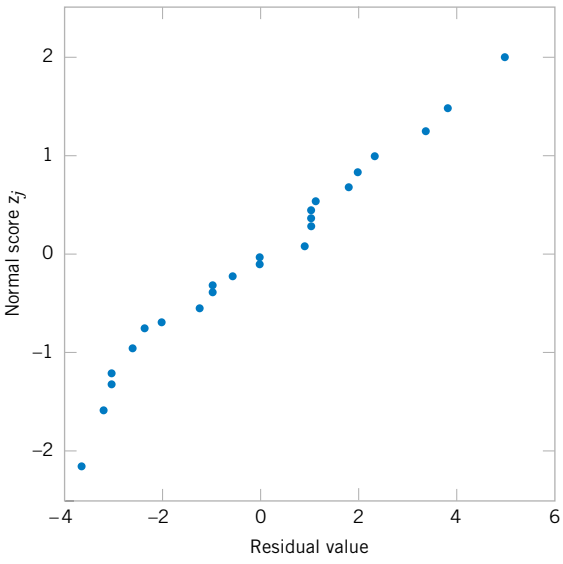


Figure 13-3 Normal probability plot of residuals from the hardwood concentration experiment.

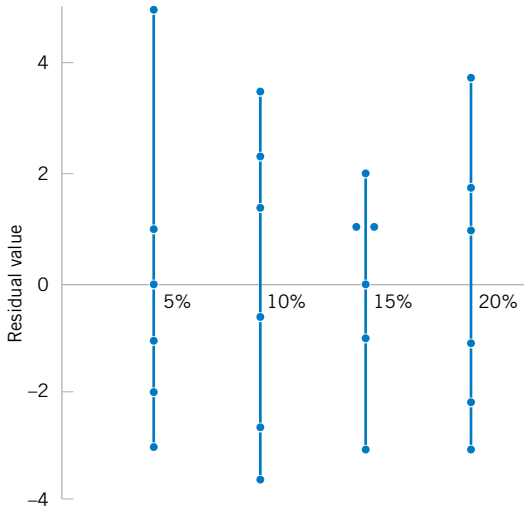


Figure 13-4 Plot of residuals versus factor levels (hardwood concentration).

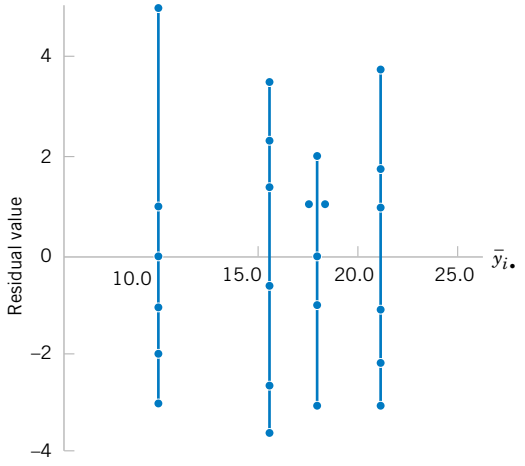


Figure 13-5 Plot of residuals versus $\bar{y}_{i.}$

This suggests that time or run order is important or that variables that change over time are important and have not been included in the experimental design.

A normal probability plot of the residuals from the paper tensile strength experiment is shown in Fig. 13-3. Figures 13-4 and 13-5 present the residuals plotted against the factor levels and the fitted value $\bar{y}_{i.}$ respectively. These plots do not reveal any model inadequacy or unusual problem with the assumptions.

13-2.6 Determining Sample Size

In any experimental design problem, the choice of the sample size or number of replicates to use is important. **Operating characteristic curves** can be used to provide guidance in making this selection. Recall that the operating characteristic curve is a plot of the probability of a

type II error (β) for various sample sizes against a measure of the difference in means that it is important to detect. Thus, if the experimenter knows the magnitude of the difference in means that is of potential importance, the operating characteristic curves can be used to determine how many replicates are required to achieve adequate sensitivity.

The power of the ANOVA test is

$$\begin{aligned} 1 - \beta &= P\{\text{Reject } H_0 \mid H_0 \text{ is false}\} \\ &= P\{F_0 > f_{\alpha, a-1, a(n-1)} \mid H_0 \text{ is false}\} \end{aligned} \tag{13-17}$$

To evaluate this probability statement, we need to know the distribution of the test statistic F_0 if the null hypothesis is false. It can be shown that, if H_0 is false, the statistic $F_0 = MS_{\text{Treatments}}/MS_E$ is distributed as a **noncentral *F* random variable,** with $a - 1$ and $a(n - 1)$ degrees of freedom and a noncentrality parameter δ. If $\delta = 0$, the noncentral F-distribution becomes the usual or *central F*-distribution.

Operating characteristic curves are used to evaluate β defined in Equation 13-17. These curves plot β against a parameter Φ, where

$$\Phi^2 = \frac{n \sum_{i=1}^{a} \tau_i^2}{a\sigma^2} \tag{13-18}$$

The parameter Φ^2 is (apart from n) the noncentrality parameter δ. Curves are available for $\alpha = 0.05$ and $\alpha = 0.01$ and for several values of the number of degrees of freedom for numerator (denoted v_1) and denominator (denoted v_2). Figure 13-6 gives representative O.C. curves, one for $a = 4$ ($v_1 = 3$) and one for $a = 5$ ($v_1 = 4$) treatments. Notice that for each value of a there are curves for $\alpha = 0.05$ and $\alpha = 0.01$. O.C. curves for other values of a are in Section 13-2.7 on the CD.

In using the operating curves, we must define the difference in means that we wish to detect in terms of $\sum_{i=1}^{a} \tau_i^2$. Also, the error variance σ^2 is usually unknown. In such cases, we must choose ratios of $\sum_{i=1}^{a} \tau_i^2/\sigma^2$ that we wish to detect. Alternatively, if an estimate of σ^2 is available, one may replace σ^2 with this estimate. For example, if we were interested in the sensitivity of an experiment that has already been performed, we might use MS_E as the estimate of σ^2.

EXAMPLE 13-3

Suppose that five means are being compared in a completely randomized experiment with $\alpha = 0.01$. The experimenter would like to know how many replicates to run if it is important to reject H_0 with probability at least 0.90 if $\sum_{i=1}^{5} \tau_i^2/\sigma^2 = 5.0$. The parameter Φ^2 is, in this case,

$$\Phi^2 = \frac{n \sum_{i=1}^{a} \tau_i^2}{a\sigma^2} = \frac{n}{5}(5) = n$$

and for the operating characteristic curve with $v_1 = a - 1 = 5 - 1 = 4$, and $v_2 = a(n - 1) = 5(n - 1)$ error degrees of freedom refer to the lower curve in Figure 13-6. As a first guess, try $n = 4$ replicates. This yields $\Phi^2 = 4$, $\Phi = 2$, and $v_2 = 5(3) = 15$ error degrees of freedom. Consequently, from Figure 13-6, we find that $\beta \simeq 0.38$. Therefore, the power of the test is approximately $1 - \beta = 1 - 0.38 = 0.62$, which is less than the required 0.90, and so we

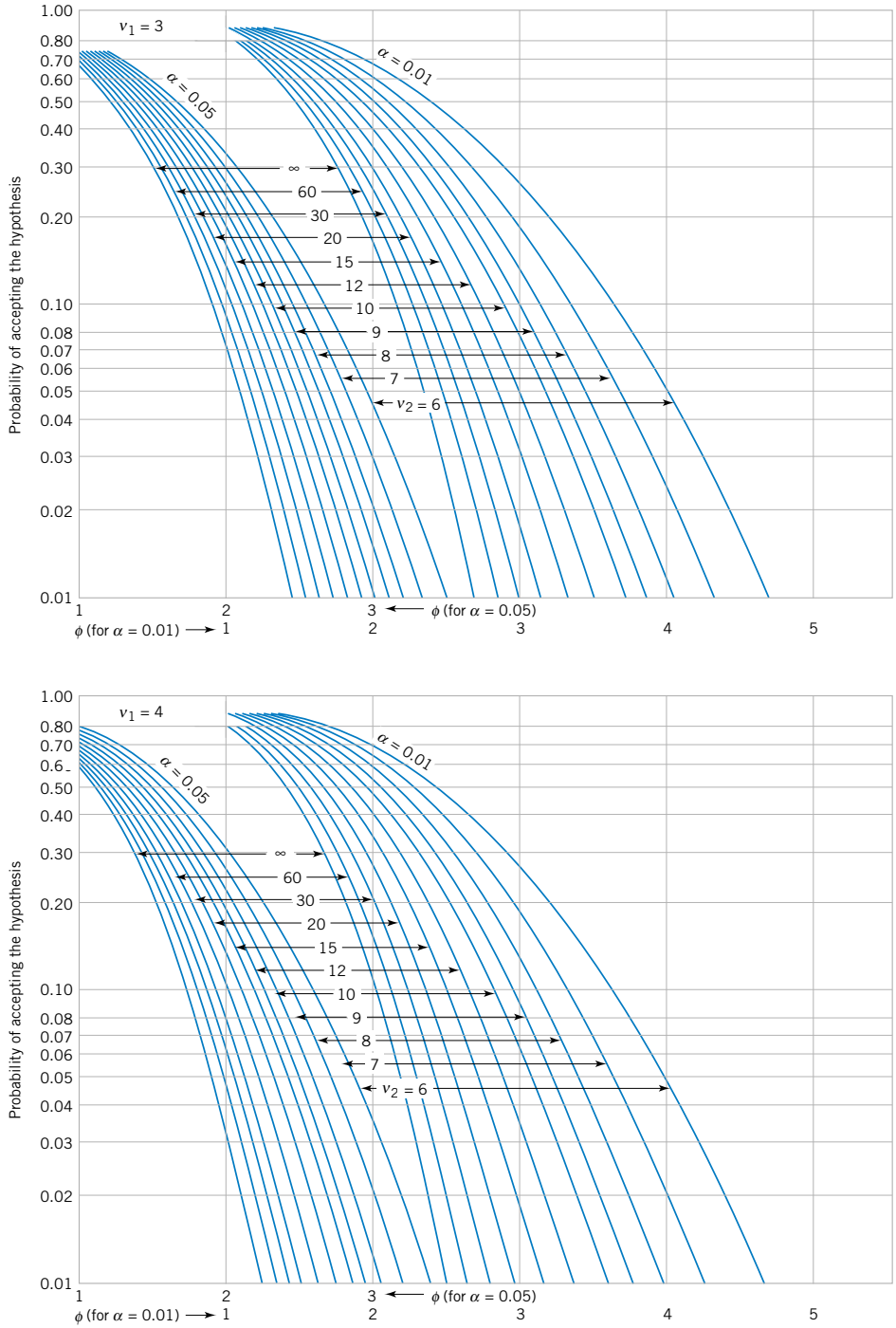


Figure 13-6 Two Operating characteristic curves for the fixed-effects model analysis of variance.

conclude that $n = 4$ replicates is not sufficient. Proceeding in a similar manner, we can construct the following table:

n	Φ^2	Φ	$a(n-1)$	β	Power $= (1-\beta)$
4	4	2.00	15	0.38	0.62
5	5	2.24	20	0.18	0.82
6	6	2.45	25	0.06	0.94

Thus, at least $n = 6$ replicates must be run in order to obtain a test with the required power.

13-2.7 Technical Details about the Analysis of Variance (CD Only)

EXERCISES FOR SECTION 13-2

13-1. In *Design and Analysis of Experiments,* 5th edition (John Wiley & Sons, 2001) D. C. Montgomery describes an experiment in which the tensile strength of a synthetic fiber is of interest to the manufacturer. It is suspected that strength is related to the percentage of cotton in the fiber. Five levels of cotton percentage are used, and five replicates are run in random order, resulting in the data below.

Cotton Percentage	Observations				
	1	2	3	4	5
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

(a) Does cotton percentage affect breaking strength? Draw comparative box plots and perform an analysis of variance. Use $\alpha = 0.05$.
(b) Plot average tensile strength against cotton percentage and interpret the results.
(c) Analyze the residuals and comment on model adequacy.

13-2. In "Orthogonal Design for Process Optimization and Its Application to Plasma Etching" (*Solid State Technology,* May 1987), G. Z. Yin and D. W. Jillie describe an experiment to determine the effect of C_2F_6 flow rate on the uniformity of the etch on a silicon wafer used in integrated circuit manufacturing. Three flow rates are used in the experiment, and the resulting uniformity (in percent) for six replicates is shown below.

C_2F_6 Flow (SCCM)	Observations					
	1	2	3	4	5	6
125	2.7	4.6	2.6	3.0	3.2	3.8
160	4.9	4.6	5.0	4.2	3.6	4.2
200	4.6	3.4	2.9	3.5	4.1	5.1

(a) Does C_2F_6 flow rate affect etch uniformity? Construct box plots to compare the factor levels and perform the analysis of variance. Use $\alpha = 0.05$.
(b) Do the residuals indicate any problems with the underlying assumptions?

13-3. The compressive strength of concrete is being studied, and four different mixing techniques are being investigated. The following data have been collected.

Mixing Technique	Compressive Strength (psi)			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

(a) Test the hypothesis that mixing techniques affect the strength of the concrete. Use $\alpha = 0.05$.
(b) Find the P-value for the F-statistic computed in part (a).
(c) Analyze the residuals from this experiment.

13-4. An experiment was run to determine whether four specific firing temperatures affect the density of a certain type of brick. The experiment led to the following data.

Temperature (°F)	Density						
100	21.8	21.9	21.7	21.6	21.7	21.5	21.8
125	21.7	21.4	21.5	21.5	—	—	—
150	21.9	21.8	21.8	21.6	21.5	—	—
175	21.9	21.7	21.8	21.7	21.6	21.8	—

(a) Does the firing temperature affect the density of the bricks? Use $\alpha = 0.05$.
(b) Find the P-value for the F-statistic computed in part (a).
(c) Analyze the residuals from the experiment.

13-5. An electronics engineer is interested in the effect on tube conductivity of five different types of coating for cathode ray tubes in a telecommunications system display device. The following conductivity data are obtained.

Coating Type	Conductivity			
1	143	141	150	146
2	152	149	137	143
3	134	133	132	127
4	129	127	132	129
5	147	148	144	142

(a) Is there any difference in conductivity due to coating type? Use $\alpha = 0.01$.
(b) Analyze the residuals from this experiment.
(c) Construct a 95% interval estimate of the coating type 1 mean. Construct a 99% interval estimate of the mean difference between coating types 1 and 4.

13-6. The response time in milliseconds was determined for three different types of circuits in an electronic calculator. The results are recorded here.

Circuit Type	Response				
1	19	22	20	18	25
2	20	21	33	27	40
3	16	15	18	26	17

(a) Using $\alpha = 0.01$, test the hypothesis that the three circuit types have the same response time.
(b) Analyze the residuals from this experiment.
(c) Find a 95% confidence interval on the response time for circuit three.

13-7. An article in the *ACI Materials Journal* (Vol. 84, 1987, pp. 213–216) describes several experiments investigating the rodding of concrete to remove entrapped air. A 3-inch × 6-inch cylinder was used, and the number of times this rod was used is the design variable. The resulting compressive strength of the concrete specimen is the response. The data are shown in the following table.

Rodding Level	Compressive Strength		
10	1530	1530	1440
15	1610	1650	1500
20	1560	1730	1530
25	1500	1490	1510

(a) Is there any difference in compressive strength due to the rodding level?
(b) Find the *P*-value for the *F*-statistic in part (a).
(c) Analyze the residuals from this experiment. What conclusions can you draw about the underlying model assumptions?

13-8. An article in *Environment International* (Vol. 18, No. 4, 1992) describes an experiment in which the amount of radon released in showers was investigated. Radon-enriched water was used in the experiment, and six different orifice diameters were tested in shower heads. The data from the experiment are shown in the following table.

Orifice Diameter	Radon Released (%)			
0.37	80	83	83	85
0.51	75	75	79	79
0.71	74	73	76	77
1.02	67	72	74	74
1.40	62	62	67	69
1.99	60	61	64	66

(a) Does the size of the orifice affect the mean percentage of radon released? Use $\alpha = 0.05$.
(b) Find the *P*-value for the *F*-statistic in part (a).
(c) Analyze the residuals from this experiment.
(d) Find a 95% confidence interval on the mean percent of radon released when the orifice diameter is 1.40.

13-9. A paper in the *Journal of the Association of Asphalt Paving Technologists* (Vol. 59, 1990) describes an experiment to determine the effect of air voids on percentage retained strength of asphalt. For purposes of the experiment, air voids are controlled at three levels; low (2–4%), medium (4–6%), and high (6–8%). The data are shown in the following table.

Air Voids	Retained Strength (%)							
Low	106	90	103	90	79	88	92	95
Medium	80	69	94	91	70	83	87	83
High	78	80	62	69	76	85	69	85

(a) Do the different levels of air voids significantly affect mean retained strength? Use $\alpha = 0.01$.
(b) Find the *P*-value for the *F*-statistic in part (a).
(c) Analyze the residuals from this experiment.
(d) Find a 95% confidence interval on mean retained strength where there is a high level of air voids.
(e) Find a 95% confidence interval on the difference in mean retained strength at the low and high levels of air voids.

13-10. An article in the *Materials Research Bulletin* (Vol. 26, No. 11, 1991) investigated four different methods of preparing the superconducting compound $PbMo_6S_8$. The authors contend

that the presence of oxygen during the preparation process affects the material's superconducting transition temperature T_c. Preparation methods 1 and 2 use techniques that are designed to eliminate the presence of oxygen, while methods 3 and 4 allow oxygen to be present. Five observations on T_c (in °K) were made for each method, and the results are as follows:

Preparation Method	Transition Temperature T_c(°K)				
1	14.8	14.8	14.7	14.8	14.9
2	14.6	15.0	14.9	14.8	14.7
3	12.7	11.6	12.4	12.7	12.1
4	14.2	14.4	14.4	12.2	11.7

(a) Is there evidence to support the claim that the presence of oxygen during preparation affects the mean transition temperature? Use $\alpha = 0.05$.
(b) What is the P-value for the F-test in part (a)?
(c) Analyze the residuals from this experiment.
(d) Find a 95% confidence interval on mean T_c when method 1 is used to prepare the material.

13-11. Use Fisher's LSD method with $\alpha = 0.05$ to analyze the means of the five different levels of cotton content in Exercise 13-1.

13-12. Use Fisher's LSD method with $\alpha = 0.05$ test to analyze the means of the three flow rates in Exercise 13-2.

13-13. Use Fisher's LSD method with $\alpha = 0.05$ to analyze the mean compressive strength of the four mixing techniques in Exercise 13-3.

13-14. Use Fisher's LSD method to analyze the five means for the coating types described in Exercise 13-5. Use $\alpha = 0.01$.

13-15. Use Fisher's LSD method to analyze the mean response times for the three circuits described in Exercise 13-6. Use $\alpha = 0.01$.

13-16. Use Fisher's LSD method to analyze the mean amounts of radon released in the experiment described in Exercise 13-8. Use $\alpha = 0.05$.

13-17. Apply Fisher's LSD method to the air void experiment described in Exercise 13-9. Using $\alpha = 0.05$, which treatment means are different?

13-18. Apply Fisher's LSD method to the superconducting material experiment described in Exercise 13-10. Which preparation methods differ, if $\alpha = 0.05$?

13-19. Suppose that four normal populations have common variance $\sigma^2 = 25$ and means $\mu_1 = 50$, $\mu_2 = 60$, $\mu_3 = 50$, and $\mu_4 = 60$. How many observations should be taken on each population so that the probability of rejecting the hypothesis of equality of means is at least 0.90? Use $\alpha = 0.05$.

13-20. Suppose that five normal populations have common variance $\sigma^2 = 100$ and means $\mu_1 = 175$, $\mu_2 = 190$, $\mu_3 = 160$, $\mu_4 = 200$, and $\mu_5 = 215$. How many observations per population must be taken so that the probability of rejecting the hypothesis of equality of means is at least 0.95? Use $\alpha = 0.01$.

13-3 THE RANDOM-EFFECTS MODEL

13-3.1 Fixed versus Random Factors

In many situations, the factor of interest has a large number of possible levels. The analyst is interested in drawing conclusions about the entire population of factor levels. If the experimenter randomly selects a of these levels from the population of factor levels, we say that the factor is a **random factor.** Because the levels of the factor actually used in the experiment were chosen randomly, the conclusions reached will be valid for the entire population of factor levels. We will assume that the population of factor levels is either of infinite size or is large enough to be considered infinite. Notice that this is a very different situation than we encountered in the fixed effects case, where the conclusions apply only for the factor levels used in the experiment.

13-3.2 ANOVA and Variance Components

The linear statistical model is

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases} \tag{13-19}$$

where the treatment effects τ_i and the errors ϵ_{ij} are independent random variables. Note that the model is identical in structure to the fixed-effects case, but the parameters have a different

interpretation. If the variance of the treatment effects τ_i is σ_τ^2, by independence the variance of the response is

$$V(Y_{ij}) = \sigma_\tau^2 + \sigma^2$$

The variances σ_τ^2 and σ^2 are called **variance components,** and the model, Equation 13-19, is called the **components of variance model** or the **random-effects model.** To test hypotheses in this model, we assume that the errors ϵ_{ij} are normally and independently distributed with mean 0 and variance σ^2 and that the treatment effects τ_i are normally and independently distributed with mean zero and variance σ_τ^2.*

For the random-effects model, testing the hypothesis that the individual treatment effects are zero is meaningless. It is more appropriate to test hypotheses about σ_τ^2. Specifically,

$$H_0: \sigma_\tau^2 = 0$$
$$H_1: \sigma_\tau^2 > 0$$

If $\sigma_\tau^2 = 0$, all treatments are identical; but if $\sigma_\tau^2 > 0$, there is variability between treatments.

The ANOVA decomposition of total variability is still valid; that is,

$$SS_T = SS_{\text{Treatments}} + SS_E \tag{13-20}$$

However, the expected values of the mean squares for treatments and error are somewhat different than in the fixed-effect case.

In the random-effects model for a single-factor, completely randomized experiment, the expected mean square for treatments is

$$E(MS_{\text{Treatments}}) = E\left(\frac{SS_{\text{Treatments}}}{a-1}\right) = \sigma^2 + n\sigma_\tau^2 \tag{13-21}$$

and the expected mean square for error is

$$E(MS_E) = E\left[\frac{SS_E}{a(n-1)}\right] = \sigma^2 \tag{13-22}$$

From examining the expected mean squares, it is clear that both MS_E and $MS_{\text{Treatments}}$ estimate σ^2 when H_0: $\sigma_\tau^2 = 0$ is true. Furthermore, MS_E and $MS_{\text{Treatments}}$ are independent. Consequently, the ratio

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E} \tag{13-23}$$

*The assumption that the $\{\tau_i\}$ are independent random variables implies that the usual assumption of $\sum_{i=1}^{a} \tau_i = 0$ from the fixed-effects model does not apply to the random-effects model.

is an F random variable with $a - 1$ and $a(n - 1)$ degrees of freedom when H_0 is true. The null hypothesis would be rejected at the α-level of significance if the computed value of the test statistic $f_0 > f_{\alpha,a-1,a(n-1)}$.

The computational procedure and construction of the ANOVA table for the random-effects model are identical to the fixed-effects case. The conclusions, however, are quite different because they apply to the entire population of treatments.

Usually, we also want to estimate the variance components (σ^2 and σ_τ^2) in the model. The procedure that we will use to estimate σ^2 and σ_τ^2 is called the **analysis of variance method** because it uses the information in the analysis of variance table. It does not require the normality assumption on the observations. The procedure consists of equating the expected mean squares to their observed values in the ANOVA table and solving for the variance components. When equating observed and expected mean squares in the one-way classification random-effects model, we obtain

$$MS_{\text{Treatments}} = \sigma^2 + n\sigma_\tau^2 \quad \text{and} \quad MS_E = \sigma^2$$

Therefore, the estimators of the variance components are

$$\hat{\sigma}^2 = MS_E \tag{13-24}$$

and

$$\hat{\sigma}_\tau^2 = \frac{MS_{\text{Treatments}} - MS_E}{n} \tag{13-25}$$

Sometimes the analysis of variance method produces a negative estimate of a variance component. Since variance components are by definition nonnegative, a negative estimate of a variance component is disturbing. One course of action is to accept the estimate and use it as evidence that the true value of the variance component is zero, assuming that sampling variation led to the negative estimate. While this approach has intuitive appeal, it will disturb the statistical properties of other estimates. Another alternative is to reestimate the negative variance component with a method that always yields nonnegative estimates. Still another possibility is to consider the negative estimate as evidence that the assumed linear model is incorrect, requiring that a study of the model and its assumptions be made to find a more appropriate model.

EXAMPLE 13-4

In *Design and Analysis of Experiments*, 5th edition (John Wiley, 2001), D. C. Montgomery describes a single-factor experiment involving the random-effects model in which a textile manufacturing company weaves a fabric on a large number of looms. The company is interested in loom-to-loom variability in tensile strength. To investigate this variability, a manufacturing engineer selects four looms at random and makes four strength determinations on fabric samples chosen at random from each loom. The data are shown in Table 13-7 and the ANOVA is summarized in Table 13-8.

From the analysis of variance, we conclude that the looms in the plant differ significantly in their ability to produce fabric of uniform strength. The variance components are estimated by $\hat{\sigma}^2 = 1.90$ and

$$\hat{\sigma}_\tau^2 = \frac{29.73 - 1.90}{4} = 6.96$$

Table 13-7 Strength Data for Example 13-4

Loom	Observations 1	2	3	4	Total	Average
1	98	97	99	96	390	97.5
2	91	90	93	92	366	91.5
3	96	95	97	95	383	95.8
4	95	96	99	98	388	97.0
					1527	95.45

Table 13-8 Analysis of Variance for the Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Looms	89.19	3	29.73	15.68	1.88 E-4
Error	22.75	12	1.90		
Total	111.94	15			

Therefore, the variance of strength in the manufacturing process is estimated by

$$\widehat{V(Y_{ij})} = \hat{\sigma}_\tau^2 + \hat{\sigma}^2 = 6.96 + 1.90 = 8.86$$

Most of this variability is attributable to differences between looms.

This example illustrates an important application of the analysis of variance—the isolation of different sources of variability in a manufacturing process. Problems of excessive variability in critical functional parameters or properties frequently arise in quality-improvement programs. For example, in the previous fabric strength example, the process mean is estimated by $\bar{y} = 95.45$ psi, and the process standard deviation is estimated by $\hat{\sigma}_y = \sqrt{\hat{V}(Y_{ij})} = \sqrt{8.86} = 2.98$ psi. If strength is approximately normally distributed, the distribution of strength in the outgoing product would look like the normal distribution shown in Fig. 13-7(a). If the lower specification limit (LSL) on strength is at 90 psi, a substantial proportion of the process output is **fallout**—that is, scrap or defective material that must be sold as second quality, and so on. This fallout is directly related to the excess variability resulting from **differences between looms.** Variability in loom performance could be caused by faulty setup, poor maintenance, inadequate supervision, poorly trained operators, and so forth. The engineer or manager responsible for quality improvement must identify and remove these sources of variability from the process. If this can be done, strength variability will be greatly reduced, perhaps as low as $\hat{\sigma}_Y = \sqrt{\hat{\sigma}^2} = \sqrt{1.90} = 1.38$ psi, as shown in Fig. 13-7(b). In this improved process, reducing the variability in strength has greatly reduced the fallout, resulting in lower cost, higher quality, a more satisfied customer, and enhanced competitive position for the company.

13-3.3 Determining Sample Size in the Random Model (CD Only)

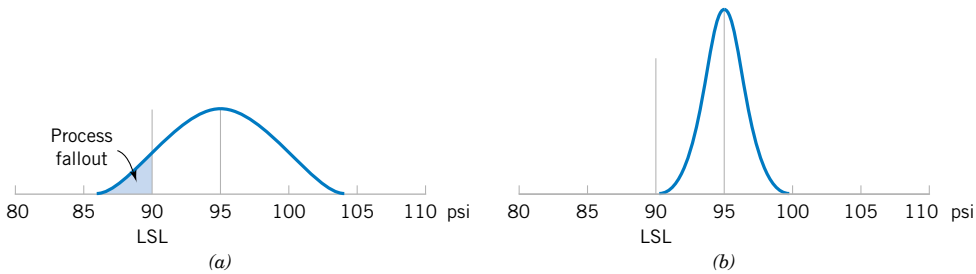


Figure 13-7 The distribution of fabric strength. (a) Current process, (b) improved process.

EXERCISES FOR SECTION 13-3

13-21. A textile mill has a large number of looms. Each loom is supposed to provide the same output of cloth per minute. To investigate this assumption, five looms are chosen at random, and their output is measured at different times. The following data are obtained:

Loom	Output (lb/min)				
1	4.0	4.1	4.2	4.0	4.1
2	3.9	3.8	3.9	4.0	4.0
3	4.1	4.2	4.1	4.0	3.9
4	3.6	3.8	4.0	3.9	3.7
5	3.8	3.6	3.9	3.8	4.0

(a) Are the looms similar in output? Use $\alpha = 0.05$.
(b) Estimate the variability between looms.
(c) Estimate the experimental error variance.
(d) Analyze the residuals from this experiment and check for model adequacy.

13-22. An article in the *Journal of the Electrochemical Society* (Vol. 139, No. 2, 1992, pp. 524–532) describes an experiment to investigate the low-pressure vapor deposition of polysilicon. The experiment was carried out in a large-capacity reactor at Sematech in Austin, Texas. The reactor has several wafer positions, and four of these positions are selected at random. The response variable is film thickness uniformity. Three replicates of the experiment were run, and the data are as follows:

Water Position	Uniformity		
1	2.76	5.67	4.49
2	1.43	1.70	2.19
3	2.34	1.97	1.47
4	0.94	1.36	1.65

(a) Is there a difference in the wafer positions? Use $\alpha = 0.05$.
(b) Estimate the variability due to wafer positions.
(c) Estimate the random error component.
(d) Analyze the residuals from this experiment and comment on model adequacy.

13-23. An article in the *Journal of Quality Technology* (Vol. 13, No. 2, 1981, pp. 111–114) describes an experiment that investigates the effects of four bleaching chemicals on pulp brightness. These four chemicals were selected at random from a large population of potential bleaching agents. The data are as follows:

Chemical	Pulp Brightness				
1	77.199	74.466	92.746	76.208	82.876
2	80.522	79.306	81.914	80.346	73.385
3	79.417	78.017	91.596	80.802	80.626
4	78.001	78.358	77.544	77.364	77.386

(a) Is there a difference in the chemical types? Use $\alpha = 0.05$.
(b) Estimate the variability due to chemical types.
(c) Estimate the variability due to random error.
(d) Analyze the residuals from this experiment and comment on model adequacy.

13-24. Consider the vapor-deposition experiment described in Exercise 13-22.
(a) Estimate the total variability in the uniformity response.
(b) How much of the total variability in the uniformity response is due to the difference between positions in the reactor?
(c) To what level could the variability in the uniformity response be reduced, if the position-to-position variability in the reactor could be eliminated? Do you believe this is a significant reduction?

13-4 RANDOMIZED COMPLETE BLOCK DESIGN

13-4.1 Design and Statistical Analysis

In many experimental design problems, it is necessary to design the experiment so that the variability arising from a nuisance factor can be controlled. For example, consider the situation of Example 10-9, where two different methods were used to predict the shear strength of steel plate girders. Because each girder has different strength (potentially), and this variability in strength was not of direct interest, we designed the experiment by using the two test methods on each girder and then comparing the average difference in strength readings on each girder to zero using the paired t-test. The paired t-test is a procedure for comparing two treatment means when all experimental runs cannot be made under

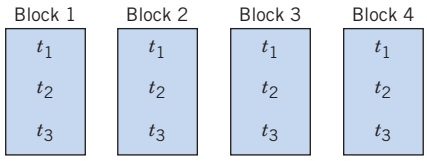


Figure 13-8 A randomized complete block design.

Table 13-9 A Randomized Complete Block Design

Treatments (Method)	Block (Girder) 1	2	3	4
1	y_{11}	y_{12}	y_{13}	y_{14}
2	y_{21}	y_{22}	y_{23}	y_{24}
3	y_{31}	y_{32}	y_{33}	y_{34}

homogeneous conditions. Alternatively, we can view the paired t-test as a method for reducing the background noise in the experiment by blocking out a **nuisance factor** effect. The block is the nuisance factor, and in this case, the nuisance factor is the actual **experimental unit**—the steel girder specimens used in the experiment.

The randomized block design is an extension of the paired t-test to situations where the factor of interest has more than two levels; that is, more than two treatments must be compared. For example, suppose that three methods could be used to evaluate the strength readings on steel plate girders. We may think of these as three treatments, say t_1, t_2, and t_3. If we use four girders as the experimental units, a **randomized complete block design** would appear as shown in Fig. 13-8. The design is called a randomized complete block design because each block is large enough to hold all the treatments and because the actual assignment of each of the three treatments within each block is done randomly. Once the experiment has been conducted, the data are recorded in a table, such as is shown in Table 13-9. The observations in this table, say y_{ij}, represent the response obtained when method i is used on girder j.

The general procedure for a randomized complete block design consists of selecting b blocks and running a complete replicate of the experiment in each block. The data that result from running a randomized complete block design for investigating a single factor with a levels and b blocks are shown in Table 13-10. There will be a observations (one per factor level) in each block, and the order in which these observations are run is randomly assigned within the block.

We will now describe the statistical analysis for a randomized complete block design. Suppose that a single factor with a levels is of interest and that the experiment is run in b blocks. The observations may be represented by the **linear statistical model**

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases} \tag{13-26}$$

Table 13-10 A Randomized Complete Block Design with a Treatments and b Blocks

Treatments	Blocks 1	2	…	b	Totals	Averages
1	y_{11}	y_{12}	$\cdots$	y_{1b}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2b}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
⋮	⋮	⋮		⋮	⋮	⋮
a	y_{a1}	y_{a2}	$\cdots$	y_{ab}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
Totals	$y_{\cdot 1}$	$y_{\cdot 2}$	$\cdots$	$y_{\cdot b}$	$y_{\cdot\cdot}$	
Averages	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	$\cdots$	$\bar{y}_{\cdot b}$		$\bar{y}_{\cdot\cdot}$

where μ is an overall mean, τ_i is the effect of the ith treatment, β_j is the effect of the jth block, and ϵ_{ij} is the random error term, which is assumed to be normally and independently distributed with mean zero and variance σ^2. Treatments and blocks will initially be considered as fixed factors. Furthermore, the treatment and block effects are defined as deviations from the overall mean, so $\sum_{i=1}^{a} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$. We also assume that treatments and blocks do not interact. That is, the effect of treatment i is the same regardless of which block (or blocks) it is tested in. We are interested in testing the equality of the treatment effects. That is

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1: \tau_i \neq 0 \text{ at least one } i$$

Testing the hypothesis that all the treatment effects τ_i are equal to zero is equivalent to testing the hypothesis that the treatment means are equal. To see this, note that the mean of the ith treatment is μ_i, defined as

$$\mu_i = E\left(\frac{\sum_{j=1}^{b} Y_{ij}}{b}\right) = \frac{1}{b}\sum_{j=1}^{b} E(Y_{ij}) = \frac{1}{b}\sum_{j=1}^{b} E(\mu + \tau_i + \beta_j + \epsilon_{ij})$$
$$= \frac{1}{b}\sum_{j=1}^{b} E(\mu + \tau_i + \beta_j) = \mu + \tau_i + \frac{1}{b}\sum_{j=1}^{b} \beta_j$$

and since $\sum_{j=1}^{b} \beta_j = 0$, we have the mean of the ith treatment defined as

$$\mu_i = \mu + \tau_i, \quad i = 1, 2, \ldots, a$$

Therefore, testing the hypothesis that the a treatment means are equal is equivalent to testing that all the treatment effects τ_i are equal to zero.

The analysis of variance can be extended to the randomized complete block design. The procedure uses a sum of squares identity that partitions the total sum of squares into three components.

The **sum of squares identity for the randomized complete block design** is

$$\sum_{i=1}^{a}\sum_{j=1}^{b} (y_{ij} - \bar{y}_{..})^2 = b\sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2 + a\sum_{j=1}^{b} (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b} (y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..})^2 \quad (13\text{-}27)$$

or symbolically

$$SS_T = SS_{\text{Treatments}} + SS_{\text{Blocks}} + SS_E$$

Furthermore, the degrees of freedom corresponding to these sums of squares are

$$ab - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1)$$

For the randomized block design, the relevant mean squares are

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1}$$

$$MS_{\text{Blocks}} = \frac{SS_{\text{Blocks}}}{b-1}$$

$$MS_E = \frac{SS_E}{(a-1)(b-1)}$$

The expected values of these mean squares can be shown to be as follows:

$$E(MS_{\text{Treatments}}) = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$E(MS_{\text{Blocks}}) = \sigma^2 + \frac{a\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$E(MS_E) = \sigma^2$$

Therefore, if the null hypothesis H_0 is true so that all treatment effects $\tau_i = 0$, $MS_{\text{Treatments}}$ is an unbiased estimator of σ^2, while if H_0 is false, $MS_{\text{Treatments}}$ overestimates σ^2. The mean square for error is always an unbiased estimate of σ^2. To test the null hypothesis that the treatment effects are all zero, we use the ratio

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E} \tag{13-28}$$

which has an F-distribution with $a-1$ and $(a-1)(b-1)$ degrees of freedom if the null hypothesis is true. We would reject the null hypothesis at the α-level of significance if the computed value of the test statistic in Equation 13-28 is $f_0 > f_{\alpha,a-1,(a-1)(b-1)}$.

In practice, we compute SS_T, $SS_{\text{Treatments}}$ and SS_{Blocks} and then obtain the error sum of squares SS_E by subtraction. The appropriate computing formulas are as follows.

Definition

The computing formulas for the sums of squares in the analysis of variance for a randomized complete block design are

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{ab} \tag{13-29}$$

$$SS_{\text{Treatments}} = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{ab} \tag{13-30}$$

$$SS_{\text{Blocks}} = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{ab} \tag{13-31}$$

and

$$SS_E = SS_T - SS_{\text{Treatments}} - SS_{\text{Blocks}} \tag{13-32}$$

Table 13-11 ANOVA for a Randomized Complete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a - 1$	$\frac{SS_{\text{Treatments}}}{a - 1}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Blocks	SS_{Blocks}	$b - 1$	$\frac{SS_{\text{Blocks}}}{b - 1}$	
Error	SS_E (by subtraction)	$(a - 1)(b - 1)$	$\frac{SS_E}{(a - 1)(b - 1)}$	
Total	SS_T	$ab - 1$		

The computations are usually arranged in an ANOVA table, such as is shown in Table 13-11. Generally, a computer software package will be used to perform the analysis of variance for the randomized complete block design.

EXAMPLE 13-5

An experiment was performed to determine the effect of four different chemicals on the strength of a fabric. These chemicals are used as part of the permanent press finishing process. Five fabric samples were selected, and a randomized complete block design was run by testing each chemical type once in random order on each fabric sample. The data are shown in Table 13-12. We will test for differences in means using an ANOVA with $\alpha = 0.01$.

The sums of squares for the analysis of variance are computed as follows:

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{5} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{ab}$$

$$= (1.3)^2 + (1.6)^2 + \cdots + (3.4)^2 - \frac{(39.2)^2}{20} = 25.69$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{4} \frac{y_{i\cdot}^2}{b} - \frac{y_{\cdot\cdot}^2}{ab}$$

$$= \frac{(5.7)^2 + (8.8)^2 + (6.9)^2 + (17.8)^2}{5} - \frac{(39.2)^2}{20} = 18.04$$

Table 13-12 Fabric Strength Data—Randomized Complete Block Design

	Fabric Sample					Treatment Totals	Treatment Averages
Chemical Type	1	2	3	4	5	$y_{i\cdot}$	$\bar{y}_{i\cdot}$
1	1.3	1.6	0.5	1.2	1.1	5.7	1.14
2	2.2	2.4	0.4	2.0	1.8	8.8	1.76
3	1.8	1.7	0.6	1.5	1.3	6.9	1.38
4	3.9	4.4	2.0	4.1	3.4	17.8	3.56
Block totals $y_{\cdot j}$	9.2	10.1	3.5	8.8	7.6	39.2($y_{\cdot\cdot}$)	
Block averages $\bar{y}_{\cdot j}$	2.30	2.53	0.88	2.20	1.90		1.96($\bar{y}_{\cdot\cdot}$)

Table 13-13 Analysis of Variance for the Randomized Complete Block Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Chemical types (treatments)	18.04	3	6.01	75.13	4.79 E-8
Fabric samples (blocks)	6.69	4	1.67		
Error	0.96	12	0.08		
Total	25.69	19			

$$SS_{\text{Blocks}} = \sum_{j=1}^{5} \frac{y_{\cdot j}^2}{a} - \frac{y_{\cdot\cdot}^2}{ab}$$
$$= \frac{(9.2)^2 + (10.1)^2 + (3.5)^2 + (8.8)^2 + (7.6)^2}{4} - \frac{(39.2)^2}{20} = 6.69$$
$$SS_E = SS_T - SS_{\text{Blocks}} - SS_{\text{Treatments}}$$
$$= 25.69 - 6.69 - 18.04 = 0.96$$

The ANOVA is summarized in Table 13-13. Since $f_0 = 75.13 > f_{0.01,3,12} = 5.95$ (the P-value is 4.79×10^{-8}), we conclude that there is a significant difference in the chemical types so far as their effect on strength is concerned.

When Is Blocking Necessary?

Suppose an experiment is conducted as a randomized block design, and blocking was not really necessary. There are ab observations and $(a - 1)(b - 1)$ degrees of freedom for error. If the experiment had been run as a completely randomized single-factor design with b replicates, we would have had $a(b - 1)$ degrees of freedom for error. Therefore, blocking has cost $a(b - 1) - (a - 1)(b - 1) = b - 1$ degrees of freedom for error. Thus, since the loss in error degrees of freedom is usually small, if there is a reasonable chance that block effects may be important, the experimenter should use the randomized block design.

For example, consider the experiment described in Example 13-5 as a single-factor experiment with no blocking. We would then have 16 degrees of freedom for error. In the randomized block design, there are 12 degrees of freedom for error. Therefore, blocking has cost only 4 degrees of freedom, which is a very small loss considering the possible gain in information that would be achieved if block effects are really important. The block effect in Example 13-5 is large, and if we had not blocked, SS_{Blocks} would have been included in the error sum of squares for the completely randomized analysis. This would have resulted in a much larger MS_E, making it more difficult to detect treatment differences. As a general rule, when in doubt as to the importance of block effects, the experimenter should block and gamble that the block effect does exist. If the experimenter is wrong, the slight loss in the degrees of freedom for error will have a negligible effect, unless the number of degrees of freedom is very small.

Computer Solution

Table 13-14 presents the computer output from Minitab for the randomized complete block design in Example 13-5. We used the analysis of variance menu for balanced designs to solve this problem. The results agree closely with the hand calculations from Table 13-13. Notice that Minitab computes an F-statistic for the blocks (the fabric samples). The validity of this ratio as a test statistic for the null hypothesis of no block effects is doubtful because the blocks represent a **restriction on randomization;** that is, we have only randomized within the blocks. If the blocks are not chosen at random, or if they are not run in random order, the

Table 13-14 Minitab Analysis of Variance for the Randomized Complete Block Design in Example 13-5

Analysis of Variance (Balanced Designs)

Factor	Type	Levels	Values				
Chemical	fixed	4	1	2	3	4	
Fabric S	fixed	5	1	2	3	4	5

Analysis of Variance for strength

Source	DF	SS	MS	F	P
Chemical	3	18.0440	6.0147	75.89	0.000
Fabric S	4	6.6930	1.6733	21.11	0.000
Error	12	0.9510	0.0792		
Total	19	25.6880			

F-test with denominator: Error
Denominator MS = 0.079250 with 12 degrees of freedom

Numerator	DF	MS	F	P
Chemical	3	6.015	75.89	0.000
Fabric S	4	1.673	21.11	0.000

F-ratio for blocks may not provide reliable information about block effects. For more discussion see Montgomery (2001, Chapter 4).

13-4.2 Multiple Comparisons

When the ANOVA indicates that a difference exists between the treatment means, we may need to perform some follow-up tests to isolate the specific differences. Any multiple comparison method, such as Fisher's LSD method, could be used for this purpose.

We will illustrate Fisher's LSD method. The four chemical type averages from Example 13-5 are:

$$\bar{y}_{1\cdot} = 1.14 \qquad \bar{y}_{2\cdot} = 1.76 \qquad \bar{y}_{3\cdot} = 1.38 \qquad \bar{y}_{4\cdot} = 3.56$$

Each treatment average uses $b = 5$ observations (one from each block). We will use $\alpha = 0.05$, so $t_{0.025,12} = 2.179$. Therefore the value of the LSD is

$$\text{LSD} = t_{0.025,12}\sqrt{\frac{2MS_E}{b}} = 2.179\sqrt{\frac{2(0.08)}{5}} = 0.39$$

Any pair of treatment averages that differ by 0.39 or more indicates that this pair of treatment means is significantly different. The comparisons are shown below:

$$\begin{aligned}
4 \text{ vs. } 1 &= \bar{y}_{4\cdot} - \bar{y}_{1\cdot} = 3.56 - 1.14 = 2.42 > 0.39 \\
4 \text{ vs. } 3 &= \bar{y}_{4\cdot} - \bar{y}_{3\cdot} = 3.56 - 1.38 = 2.18 > 0.39 \\
4 \text{ vs. } 2 &= \bar{y}_{4\cdot} - \bar{y}_{2\cdot} = 3.56 - 1.76 = 1.80 > 0.39 \\
2 \text{ vs. } 1 &= \bar{y}_{2\cdot} - \bar{y}_{1\cdot} = 1.76 - 1.14 = 0.62 > 0.39 \\
2 \text{ vs. } 3 &= \bar{y}_{2\cdot} - \bar{y}_{3\cdot} = 1.76 - 1.38 = 0.38 < 0.39 \\
3 \text{ vs. } 1 &= \bar{y}_{3\cdot} - \bar{y}_{1\cdot} = 1.38 - 1.14 = 0.24 < 0.39
\end{aligned}$$

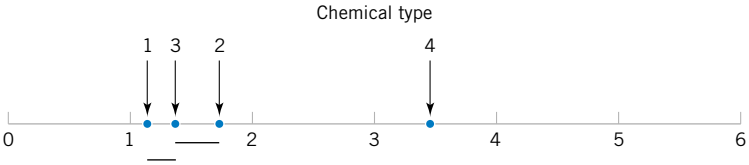


Figure 13-9 Results of Fisher's LSD method.

Figure 13-9 presents the results graphically. The underlined pairs of means are not different. The LSD procedure indicates that chemical type 4 results in significantly different strengths than the other three types do. Chemical types 2 and 3 do not differ, and types 1 and 3 do not differ. There may be a small difference in strength between types 1 and 2.

13-4.3 Residual Analysis and Model Checking

In any designed experiment, it is always important to examine the residuals and to check for violation of basic assumptions that could invalidate the results. As usual, the residuals for the randomized complete block design are just the difference between the observed and estimated (or fitted) values from the statistical model, say,

$$e_{ij} = y_{ij} - \hat{y}_{ij} \tag{13-33}$$

and the fitted values are

$$\hat{y}_{ij} = \bar{y}_{i\cdot} + \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}$$

The fitted value represents the estimate of the mean response when the ith treatment is run in the jth block. The residuals from the chemical type experiment are shown in Table 13-15.

Figures 13-10, 13-11, 13-12, and 13-13 present the important residual plots for the experiment. These residual plots are usually constructed by computer software packages. There is some indication that fabric sample (block) 3 has greater variability in strength when treated with the four chemicals than the other samples. Chemical type 4, which provides the greatest strength, also has somewhat more variability in strength. Followup experiments may be necessary to confirm these findings, if they are potentially important.

13-4.4 Randomized Complete Block Design with Random Factors (CD Only)

Table 13-15 Residuals from the Randomized Complete Block Design

Chemical Type	Fabric Sample				
	1	2	3	4	5
1	−0.18	−0.10	0.44	−0.18	0.02
2	0.10	0.08	−0.28	0.00	0.10
3	0.08	−0.24	0.30	−0.12	−0.02
4	0.00	0.28	−0.48	0.30	−0.10

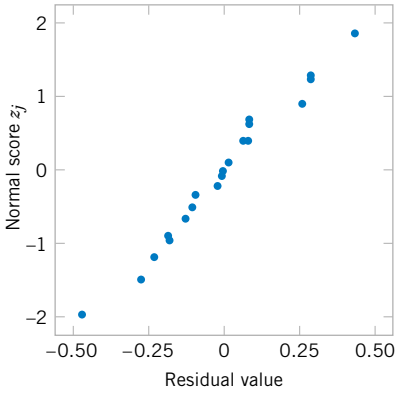


Figure 13-10 Normal probability plot of residuals from the randomized complete block design.

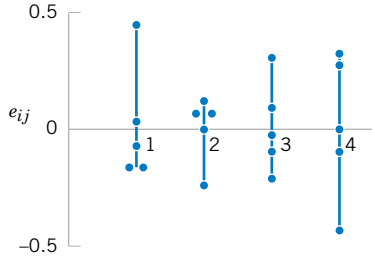


Figure 13-11 Residuals by treatment.

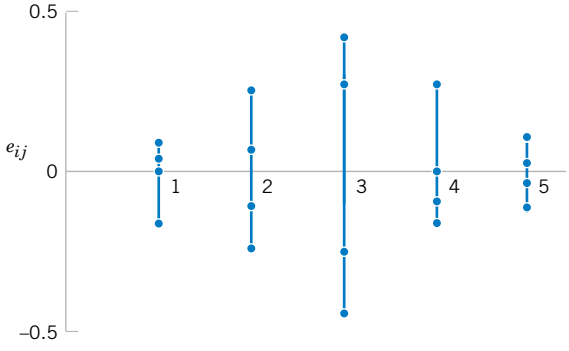


Figure 13-12 Residuals by block.

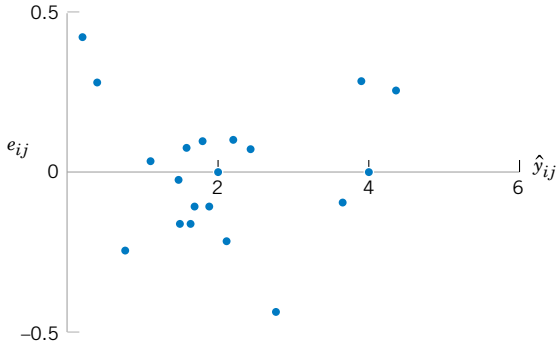


Figure 13-13 Residuals versus $\hat{y}_{ij}$.

EXERCISES FOR SECTION 13-4

13-25. In "The Effect of Nozzle Design on the Stability and Performance of Turbulent Water Jets" (*Fire Safety Journal,* Vol. 4, August 1981), C. Theobald describes an experiment in which a shape measurement was determined for several different nozzle types at different levels of jet efflux velocity. Interest in this experiment focuses primarily on nozzle type, and velocity is a nuisance factor. The data are as follows:

Nozzle Type	Jet Efflux Velocity (m/s)					
	11.73	14.37	16.59	20.43	23.46	28.74
1	0.78	0.80	0.81	0.75	0.77	0.78
2	0.85	0.85	0.92	0.86	0.81	0.83
3	0.93	0.92	0.95	0.89	0.89	0.83
4	1.14	0.97	0.98	0.88	0.86	0.83
5	0.97	0.86	0.78	0.76	0.76	0.75

(a) Does nozzle type affect shape measurement? Compare the nozzles with box plots and the analysis of variance.

(b) Use Fisher's LSD method to determine specific differences between the nozzles. Does a graph of the average (or standard deviation) of the shape measurements versus nozzle type assist with the conclusions?

(c) Analyze the residuals from this experiment.

13-26. In *Design and Analysis of Experiments,* 5th edition (John Wiley & Sons, 2001), D. C. Montgomery describes an experiment that determined the effect of four different types of tips in a hardness tester on the observed hardness of a metal alloy. Four specimens of the alloy were obtained, and each tip was tested once on each specimen, producing the following data:

(a) Is there any difference in hardness measurements between the tips?

Type of Tip	Specimen			
	1	2	3	4
1	9.3	9.4	9.6	10.0
2	9.4	9.3	9.8	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

(b) Use Fisher's LSD method to investigate specific differences between the tips.
(c) Analyze the residuals from this experiment.

13-27. An article in the *American Industrial Hygiene Association Journal* (Vol. 37, 1976, pp. 418–422) describes a field test for detecting the presence of arsenic in urine samples. The test has been proposed for use among forestry workers because of the increasing use of organic arsenics in that industry. The experiment compared the test as performed by both a trainee and an experienced trainer to an analysis at a remote laboratory. Four subjects were selected for testing and are considered as blocks. The response variable is arsenic content (in ppm) in the subject's urine. The data are as follows:

Test	Subject			
	1	2	3	4
Trainee	0.05	0.05	0.04	0.15
Trainer	0.05	0.05	0.04	0.17
Lab	0.04	0.04	0.03	0.10

(a) Is there any difference in the arsenic test procedure?
(b) Analyze the residuals from this experiment.

13-28. An article in the *Food Technology Journal* (Vol. 10, 1956, pp. 39–42) describes a study on the protopectin content of tomatoes during storage. Four storage times were selected, and samples from nine lots of tomatoes were analyzed. The protopectin content (expressed as hydrochloric acid soluble fraction mg/kg) is in the following table.

(a) The researchers in this study hypothesized that mean protopectin content would be different at different storage times. Can you confirm this hypothesis with a statistical test using $\alpha = 0.05$?
(b) Find the *P*-value for the test in part (a).
(c) Which specific storage times are different? Would you agree with the statement that protopectin content decreases as storage time increases?
(d) Analyze the residuals from this experiment.

13-29. An experiment was conducted to investigate leaking current in a SOS MOSFETS device. The purpose of the experiment was to investigate how leakage current varies as the channel length changes. Four channel lengths were selected. For each channel length, five different widths were also used, and width is to be considered a nuisance factor. The data are as follows:

Channel Length	Width				
	1	2	3	4	5
1	0.7	0.8	0.8	0.9	1.0
2	0.8	0.8	0.9	0.9	1.0
3	0.9	1.0	1.7	2.0	4.0
4	1.0	1.5	2.0	3.0	20.0

(a) Test the hypothesis that mean leakage voltage does not depend on the channel length, using $\alpha = 0.05$.
(b) Analyze the residuals from this experiment. Comment on the residual plots.

13-30. Consider the leakage voltage experiment described in Exercise 13-29. The observed leakage voltage for channel length 4 and width 5 was erroneously recorded. The correct observation is 4.0. Analyze the corrected data from this experiment. Is there evidence to conclude that mean leakage voltage increases with channel length?

Supplemental Exercises

13-31. An article in the *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* (Vol. 15, No. 2, 1992, pp. 146–153) describes an experiment in which the contact resistance of a brake-only relay was studied

Storage Time	Lot								
	1	2	3	4	5	6	7	8	9
0 days	1694.0	989.0	917.3	346.1	1260.0	965.6	1123.0	1106.0	1116.0
7 days	1802.0	1074.0	278.8	1375.0	544.0	672.2	818.0	406.8	461.6
14 days	1568.0	646.2	1820.0	1150.0	983.7	395.3	422.3	420.0	409.5
21 days	415.5	845.4	377.6	279.4	447.8	272.1	394.1	356.4	351.2

for three different materials (all were silver-based alloys). The data are as follows.

Alloy	Contact Resistance				
1	95	97	99	98	99
	99	99	94	95	98
2	104	102	102	105	99
	102	111	103	100	103
3	119	130	132	136	141
	172	145	150	144	135

(a) Does the type of alloy affect mean contact resistance? Use $\alpha = 0.01$.
(b) Use Fisher's LSD method to determine which means differ.
(c) Find a 99% confidence interval on the mean contact resistance for alloy 3.
(d) Analyze the residuals for this experiment.

13-32. An article in *Lubrication Engineering* (December 1990) describes the results of an experiment designed to investigate the effects of carbon material properties on the progression of blisters on carbon face seals. The carbon face seals are used extensively in equipment such as air turbine starters. Five different carbon materials were tested, and the surface roughness was measured. The data are as follows:

Carbon Material Type	Surface Roughness					
EC10	0.50	0.55	0.55	0.36		
EC10A	0.31	0.07	0.25	0.18	0.56	0.20
EC4	0.20	0.28	0.12			
EC1	0.10	0.16				

(a) Does carbon material type have an effect on mean surface roughness? Use $\alpha = 0.05$.
(b) Find the residuals for this experiment. Does a normal probability plot of the residuals indicate any problem with the normality assumption?
(c) Plot the residuals versus $\hat{y}_{ij}$. Comment on the plot.
(d) Find a 95% confidence interval on the difference between the mean surface roughness between the EC10 and the EC1 carbon grades.

13-33. Apply the Fisher LSD method to the experiment in Exercise 13-32. Summarize your conclusions regarding the effect of material type on surface roughness.

13-34. An article in the *Journal of Quality Technology* (Vol. 14, No. 2, 1982, pp. 80–89) describes an experiment in which three different methods of preparing fish are evaluated on the basis of sensory criteria and a quality score is assigned. Assume that these methods have been randomly selected from a large population of preparation methods. The data are in the following table:

Method	Score			
1	24.4	23.2	25.0	19.7
	22.2	24.4	23.8	18.0
2	22.1	19.5	17.3	19.7
	22.3	23.2	21.4	22.6
3	23.3	22.8	22.4	23.7
	20.4	23.5	20.8	24.1

(a) Is there any difference in preparation methods? Use $\alpha = 0.05$.
(b) Calculate the P-value for the F-statistic in part (a).
(c) Analyze the residuals from this experiment and comment on model adequacy.
(d) Estimate the components of variance.

13-35. An article in the *Journal of Agricultural Engineering Research* (Vol. 52, 1992, pp. 53–76) describes an experiment to investigate the effect of drying temperature of wheat grain on the baking quality of bread. Three temperature levels were used, and the response variable measured was the volume of the loaf of bread produced. The data are as follows:

Temperature (°C)	Volume (CC)				
70.0	1245	1235	1285	1245	1235
75.0	1235	1240	1200	1220	1210
80.0	1225	1200	1170	1155	1095

(a) Does drying temperature affect mean bread volume? Use $\alpha = 0.01$.
(b) Find the P-value for this test.
(c) Use the Fisher's LSD method to determine which means are different.
(d) Analyze the residuals from this experiment and comment on model adequacy.

13-36. An article in *Agricultural Engineering* (December 1964, pp. 672–673) describes an experiment in which the daily weight gain of swine is evaluated at different levels of housing temperature. The mean weight of each group of swine at the start of the experiment is considered to be a nuisance factor. The data from this experiment are as follows:

Mean Weight (lbs)	Housing Air Temperatures (degrees F)					
	50	60	70	80	90	100
100	1.37	1.58	2.00	1.97	1.40	0.39
150	1.47	1.75	2.16	1.82	1.14	−0.19
200	1.19	1.91	2.22	1.67	0.88	−0.77

(a) Does housing air temperature affect mean weight gain? Use $\alpha = 0.05$.
(b) Use Fisher's LSD method to determine which temperature levels are different.
(c) Analyze the residuals from this experiment and comment on model adequacy.

13-37. An article in *Communications of the ACM* (Vol. 30, No. 5, 1987) studied different algorithms for estimating software development costs. Six algorithms were applied to eight software development projects and the percent error in estimating the development cost was observed. The data are in the table at the bottom of the page.

(a) Do the algorithms differ in their mean cost estimation accuracy? Use $\alpha = 0.05$.
(b) Analyze the residuals from this experiment.
(c) Which algorithm would you recommend for use in practice?

13-38. Consider an ANOVA situation with $a = 4$ means $\mu_1 = 1$, $\mu_2 = 5$, $\mu_3 = 8$, and $\mu_4 = 4$. Suppose that $\sigma^2 = 4$, $n = 4$, and $\alpha = 0.05$.
(a) Find the power of the ANOVA F-test.
(b) How large would the sample size have to be if we want the power of the F-test for detecting this difference in means to be at least 0.90?

13-39. Consider an ANOVA situation with $a = 5$ treatments. Let $\sigma^2 = 9$ and $\alpha = 0.05$, and suppose that $n = 4$.
(a) Find the power of the ANOVA F-test when $\mu_1 = \mu_2 = \mu_3 = 1$, $\mu_4 = 3$, and $\mu_5 = 2$.
(b) What sample size is required if we want the power of the F-test in this situation to be at least 0.90?

Algorithm	Project							
	1	2	3	4	5	6	7	8
1(SLIM)	1244	21	82	2221	905	839	527	122
2(COCOMO-A)	281	129	396	1306	336	910	473	199
3(COCOMO-R)	220	84	458	543	300	794	488	142
4(COCOMO-C)	225	83	425	552	291	826	509	153
5(FUNCTION POINTS)	19	11	−34	121	15	103	87	−17
6(ESTIMALS)	−20	35	−53	170	104	199	142	41

MIND-EXPANDING EXERCISES

13-40. Show that in the fixed-effects model analysis of variance $E(MS_E) = \sigma^2$. How would your development change if the random-effects model had been specified?

13-41. Consider testing the equality of the means of two normal populations where the variances are unknown but are assumed equal. The appropriate test procedure is the two-sample t-test. Show that the two-sample t-test is equivalent to the single-factor analysis of variance F-test.

13-42. Consider the ANOVA with $a = 2$ treatments. Show that the MS_E in this analysis is equal to the pooled variance estimate used in the two-sample t-test.

13-43. Show that the variance of the linear combination

$$\sum_{i=1}^{a} c_i Y_{i\cdot} \text{ is } \sigma^2 \sum_{i=1}^{a} n_i c_i^2.$$

13-44. In a fixed-effects model, suppose that there are n observations for each of four treatments. Let Q_1^2, Q_2^2, and Q_3^2 be single-degree-of-freedom sums of squares for the orthogonal contrasts. Prove that $SS_{\text{Treatments}} = Q_1^2 + Q_2^2 + Q_3^2$.

13-45. Consider the single-factor completely randomized design with a treatments and n replicates. Show that if the difference between any two treatment means is as large as D, the minimum value that the OC curve parameter Φ^2 can take on is

$$\Phi^2 = \frac{nD^2}{2a\sigma^2}$$

13-46. Consider the single-factor completely randomized design. Show that a $100(1 - \alpha)$ percent confidence interval for σ^2 is

$$\frac{(N-a)MS_E}{\chi^2_{\alpha/2,N-a}} \le \sigma^2 \le \frac{(N-a)MS_E}{\chi^2_{1-\alpha/2,N-a}}$$

where N is the total number of observations in the experimental design.

13-47. Consider the random-effect model for the single-factor completely randomized design. Show that a $100(1 - \alpha)\%$ confidence interval on the ratio of variance components σ_τ^2/σ^2 is given by

$$L \le \frac{\sigma_\tau^2}{\sigma^2} \le U$$

where

$$L = \frac{1}{n}\left[\frac{MS_{\text{Treatments}}}{MS_E} \times \left(\frac{1}{f_{\alpha/2,a-1,N-a}}\right) - 1\right]$$

and

$$U = \frac{1}{n}\left[\frac{MS_{\text{Treatments}}}{MS_E} \times \left(\frac{1}{f_{1-\alpha/2,a-1,N-a}}\right) - 1\right]$$

13-48. Consider a random-effects model for the single-factor completely randomized design. Show that a $100(1 - \alpha)\%$ confidence interval on the ratio $\sigma_\tau^2/(\sigma^2 + \sigma_\tau^2)$ is

$$\frac{L}{1+L} \le \frac{\sigma_\tau^2}{\sigma^2 + \sigma_\tau^2} \le \frac{U}{1+U}$$

where L and U are as defined in Exercise 13-47.

13-49. Continuation of Exercise 13-48. Use the results of Exercise 13-48 to find a $100(1 - \alpha)\%$ confidence interval for $\sigma^2/(\sigma^2 + \sigma_\tau^2)$.

13-50. Consider the fixed-effect model of the completely randomized single-factor design. The model parameters are restricted by the constraint $\sum_{i=1}^{a} \tau_i = 0$. (Actually, other restrictions could be used, but this one is simple and results in intuitively pleasing estimates for the model parameters.) For the case of unequal sample size $n_1, n_2, \ldots, n_a$, the restriction is $\sum_{i=1}^{a} n_i\tau_i = 0$. Use this to show that

$$E(MS_{\text{Treatments}}) = \sigma^2 + \frac{\sum_{i=1}^{a} n_i\tau_i^2}{a-1}$$

Does this suggest that the null hypothesis in this model is H_0: $n_1\tau_1 = n_2\tau_2 = \cdots = n_a\tau_a = 0$?

13-51. Sample Size Determination. In the single-factor completely randomized design, the accuracy of a

MIND-EXPANDING EXERCISES

100(1 − α)% confidence interval on the difference in any two treatment means is $t_{\alpha/2,a(n-1)}\sqrt{2MS_E/n}$.

(a) Show that if A is the desired accuracy of the interval, the sample size required is

$$n = \frac{2F_{\alpha/2,1,a(n-1)}MS_E}{A^2}$$

(b) Suppose that in comparing $a = 5$ means we have a preliminary estimate of σ^2 of 4. If we want the 95% confidence interval on the difference in means to have an accuracy of 2, how many replicates should we use?

IMPORTANT TERMS AND CONCEPTS

In the E-book, click on any term or concept below to go to that subject.

Analysis of variance
Blocking
Complete randomized experiment
Expected mean squares
Fisher's least significant difference method
Fixed factor
Multiple comparisons
Nuisance factors
Random factor
Randomization
Randomized complete block design
Residual analysis and model adequacy checking
Sample size and replication in an experiment
Variance component

CD MATERIAL

Graphical comparison of means
Orthogonal contrasts
Tukey's test

13-2.4 More About Multiple Comparisons

As noted in the previous section, there are many ways to investigate the treatment means following rejection of the null hypothesis with an ANOVA. The Fisher LSD method is easy and very widely used. It is consider to be a very "liberal" procedure in that although each test is at significance level α, the type I error for the entire set of comparisons (called the experimentwise error rate) is much greater than α. In this section we briefly describe three other approaches.

Graphical Comparison of Means

It is easy to compare treatment means graphically, following the analysis of variance. Suppose that the factor has a levels and that $\bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \ldots, \bar{y}_{a\cdot}$ are the observed averages for these factor levels. Each treatment average has standard deviation $\sigma/\sqrt{n}$, where σ is the standard deviation of an individual observation. If all treatment means are equal, the observed means $\bar{y}_{i\cdot}$ would behave as if they were a set of observations drawn at random from a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

Visualize this normal distribution capable of being slid along an axis below which the treatment means $\bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \ldots, \bar{y}_{a\cdot}$ are plotted. If all treatment means are equal, there should be some position for this distribution that makes it obvious that the $\bar{y}_{i\cdot}$ values were drawn from the same distribution. If this is not the case, the $\bar{y}_{i\cdot}$ values that do not appear to have been drawn from this distribution are associated with treatments that produce different mean responses.

The only flaw in this logic is that σ is unknown. However, we can use $\sqrt{MS_E}$ from the analysis of variance to estimate σ. This implies that a t-distribution should be used instead of the normal in making the plot, but since the t looks so much like the normal, sketching a normal curve that is approximately $6\sqrt{MS_E/n}$ units wide will usually work very well.

Figure S13-1 shows this arrangement for the hardwood concentration experiment in Example 13-1. The standard deviation of this normal distribution is

$$\sqrt{MS_E/n} = \sqrt{6.51/6} = 1.04$$

If we visualize sliding this distribution along the horizontal axis, we note that there is no location for the distribution that would suggest that all four observations (the plotted means) are typical, randomly selected values from that distribution. This, of course, should be expected, because the analysis of variance has indicated that the means differ, and the display in Fig. S13-1 is just a graphical representation of the analysis of variance results. The figure does

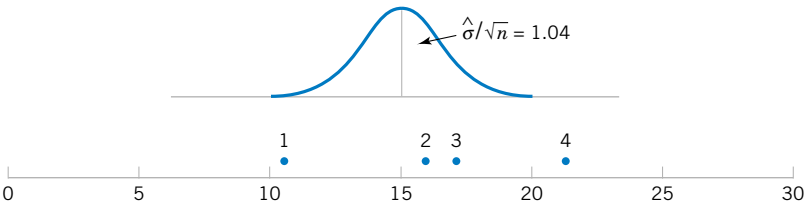


Figure S13-1 Tensile strength averages from the hardwood concentration experiment in relation to a normal distribution with standard deviation $\sqrt{MS_E/n} = \sqrt{6.51/6} = 1.04$.

indicate that treatment 4 (20% hardwood) produces paper with higher mean tensile strength than do the other treatments, and treatment 1 (5% hardwood) results in lower mean tensile strength than do the other treatments. The means of treatments 2 and 3 (10 and 15% hardwood, respectively) do not differ.

This simple procedure is a rough but very effective multiple comparison technique. We now briefly describe two other procedures: orthogonal contrasts and Tukey's method.

Orthogonal Contrasts

Many multiple comparison procedures use the idea of a contrast. Consider the hardwood concentration experiment presented in Example 13-1. Since the hypothesis H_0: $\tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$ was rejected, we know that some hardwood concentrations produce different tensile strengths than others, but which ones actually cause this difference? At the outset of the experiment, we might suspect that hardwood concentrations 3 and 4 produce the same tensile strength. This implies that we would like to test the hypothesis

$$H_0: \mu_3 = \mu_4$$
$$H_1: \mu_3 \neq \mu_4$$

This hypothesis could be tested by using a linear combination of treatment totals, say,

$$y_{3\cdot} - y_{4\cdot}$$

If we had suspected that the *average* of hardwood concentrations 1 and 3 did not differ from the average of hardwood concentrations 2 and 4, the hypothesis would have been

$$H_0: \mu_1 + \mu_3 = \mu_2 + \mu_4$$
$$H_1: \mu_1 + \mu_3 \neq \mu_2 + \mu_4$$

which implies using the linear combination of treatment totals

$$y_{1\cdot} + y_{3\cdot} - y_{2\cdot} - y_{4\cdot}$$

In general, the comparison of treatment means of interest will imply a linear combination of treatment totals such as

$$c = \sum_{i=1}^{a} c_i y_{i\cdot}$$

with the restriction that $\sum_{i=1}^{a} c_i = 0$. These linear combinations are called **contrasts.** The sum of squares for any contrast is

$$SS_c = \frac{\left(\sum_{i=1}^{a} c_i y_{i\cdot}\right)^2}{n\sum_{i=1}^{a} c_i^2} \tag{S13-1}$$

and has a single degree of freedom. If the design is unbalanced, the comparison of treatment means requires that $\sum_{i=1}^{a} n_i c_i = 0$, and Equation S13-1 becomes

$$SS_c = \frac{\left(\sum_{i=1}^{a} c_i y_{i\cdot}\right)^2}{\sum_{i=1}^{a} n_i c_i^2} \tag{S13-2}$$

A contrast is tested by comparing its sum of squares to the mean square error. The resulting statistic is distributed as F, with 1 and $N - a$ degrees of freedom.

A very important special case of the above procedure is that of **orthogonal contrasts.** Two contrasts with coefficients $\{c_i\}$ and $\{d_i\}$ are orthogonal if

$$\sum_{i=1}^{a} c_i d_i = 0$$

or for an unbalanced design if

$$\sum_{i=1}^{a} n_i c_i d_i = 0$$

For a treatments a set of $a - 1$ orthogonal contrasts will partition the sum of squares due to treatments into $a - 1$ independent single-degree-of-freedom sums of squares. Thus, tests performed on orthogonal contrasts are independent.

There are many ways to choose the orthogonal contrast coefficients for a set of treatments. Usually, something in the context of the experiment should suggest which comparisons will be of interest. For example, if there are $a = 3$ treatments, with treatment 1 a control and treatments 2 and 3 actual levels of the factor of interest to the experimenter, appropriate orthogonal contrasts might be as follows:

$$H_0: -2\mu_1 + \mu_2 + \mu_3 = 0$$
$$H_0: \qquad \mu_2 - \mu_3 = 0$$

Note that contrast 1 with $c_i = -2, 1, 1$ compares the average effect of the factor with the control, while contrast 2 with $d_i = 0, 1, -1$ compares the two levels of the factor of interest.

Contrast coefficients must be chosen prior to running the experiment, because if these comparisons are selected after examining the data, most experimenters would construct tests that compare large observed differences in means. These large differences could be due to the presence of real effects, or they could be due to random error. If experimenters always pick the largest differences to compare, they will inflate the type I error of the test, since it is likely that in an unusually high percentage of the comparisons selected, the observed differences will be due to error.

EXAMPLE S13-1 Consider the hardwood concentration experiment. There are four levels of hardwood concentration, and possible sets of comparisons between these means and the associated orthogonal comparisons are

$$H_0: \mu_1 + \mu_4 = \mu_2 + \mu_3 \qquad c = \quad y_{1\cdot} - y_{2\cdot} - y_{3\cdot} + y_{4\cdot}$$
$$H_0: \mu_1 + \mu_2 = \mu_3 + \mu_4 \qquad d = -y_{1\cdot} - y_{2\cdot} + y_{3\cdot} + y_{4\cdot}$$
$$H_0: \mu_1 + \mu_3 = \mu_2 + \mu_4 \qquad e = -y_{1\cdot} + y_{2\cdot} - y_{3\cdot} + y_{4\cdot}$$

Table S13-1 Analysis of Variance for the Tensile Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0
Hardwood concentration	382.79	3	127.60	19.61
c (1, 4 vs. 2, 3)	3.38	1	3.38	0.52
d (1, 2 vs. 3, 4)	234.38	1	234.38	36.00
e (1, 3 vs. 2, 4)	145.04	1	145.04	22.28
Error	130.17	20	6.51	
Total	512.96	23		

Notice that the contrast constants are orthogonal. Using the data from Table S13-1, we find the numerical values of the contrasts and the sums of squares as follows:

$$c = 60 - 94 - 102 + 127 = -9 \qquad SS_c = \frac{(-9)^2}{6(4)} = 3.38$$

$$d = -60 - 94 + 102 + 127 = 75 \qquad SS_d = \frac{(75)^2}{6(4)} = 234.38$$

$$e = -60 + 94 - 102 + 127 = 59 \qquad SS_e = \frac{(59)^2}{6(4)} = 145.04$$

These contrast sums of squares completely partition the treatment sum of squares; that is, $SS_{\text{Treatments}} = SS_c + SS_d + SS_e$. These tests on the contrasts are usually incorporated in the analysis of variance, such as is shown in Table S13-1. From this analysis, we conclude that there are significant differences between hardwood concentration 3 and 4, and 1 and 2, but that the average of 1 and 4 does not differ from the average of 2 and 3. Also, the average of 1 and 3 differs from the average of 2 and 4.

Tukey's Method

The Tukey procedure for comparing pairs of means makes use of the **studentized range statistic**

$$Q = \frac{\bar{Y}_{\max} - \bar{Y}_{\min}}{\sqrt{MS_E/n}}$$

where $\bar{Y}_{\max}$ and $\bar{Y}_{\min}$ are the largest and smallest sample means, respectively, out of a group of p sample means. For equal sample sizes, the Tukey procedure would indicate that the two means μ_i and μ_j are different if the absolute value of the observed difference $|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}|$ exceeds

$$T_\alpha = g_\alpha(a, f)\sqrt{\frac{MS_E}{n}}$$

where $g_\alpha(a, f)$ is the upper α percentage point of the studentized range statistic, a is the number of treatments, and f is the number of even degrees of freedom. Tables of $g_\alpha(a, f)$ are

Table S13-2 Minitab Output Illustrating Tukey's Method

One-way ANOVA: Tensile Str versus Conc

Analysis of Variance for Tensile

Source	DF	SS	MS	F	P
Conc	3	382.79	127.60	19.61	0.000
Error	20	130.17	6.51		
Total	23	512.96			

Level	N	Mean	StDev
5	6	10.000	2.828
10	6	15.667	2.805
15	6	17.000	1.789
20	6	21.167	2.639

Pooled StDev = 2.551

Individual 95% CIs For Mean Based on Pooled StDev

```
-----+----------+----------+----------+-
(---*---)
          (---*----)
          (---*----)
                    (---*----)
-----+----------+----------+----------+-
10.0      15.0       20.0       25.0
```

Tukey's pairwise comparisons

Family error rate = 0.0500
Individual error rate = 0.0111

Critical value = 3.96

Intervals for (column level mean) − (row level mean)

	5	10	15
10	−9.791 −1.542		
15	−11.124 −2.876	−5.458 2.791	
20	−15.291 −7.042	−9.624 −1.376	−8.291 −0.042

widely available; for example, see Montgomery (2001). Equivalently, we could construct a set of $100(1 - \alpha)\%$ confidence intervals for all pairs of mean using

$$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} - T_\alpha \le \mu_i - \mu_j \le \bar{y}_{i\cdot} - \bar{y}_{j\cdot} + T_\alpha$$

For unequal sample sizes, use

$$T_\alpha = \frac{g_\alpha(a, f)}{\sqrt{2}} \sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

The Tukey confidence intervals are a set of simultaneous confidence intervals that hold with probability $1 - \alpha$. Tukey's method is a very conservative procedure relative to Fisher's LSD because it requires a larger observed difference in treatment averages to declair the pair of means different.

Minitab implements the Tukey procedure and reports the results in terms of the confidence interval. Table S13-2 is the Minitab output for the hardwood concentration experiment of Example S13-1. Notice that, like Fisher's LSD, Tukey's method indicates that all pairs of means are different except at 10% and 15% concentrations.

13-2.7 Technical Details about the Analysis of Variance (CD Only)

Derivation of the ANOVA Identity

The proof of the fundamental ANOVA identity in Equation 13-5 is straightforward. Note that we may write

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2=\sum_{i=1}^{a}\sum_{j=1}^{n}[(\bar{y}_{i.}-\bar{y}_{..})+(y_{ij}-\bar{y}_{i.})]^2$$

or

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2=n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2+2\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})(y_{ij}-\bar{y}_{i.})$$

Note that the cross-product term in the previous equation is zero, since

$$\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})=y_{i.}-n\bar{y}_{i.}=y_{i.}-n(y_{i.}/n)=0$$

Therefore, we have shown that Equation 13-5 is correct.

Expected Mean Squares

In the text we state that

$$E(MS_{\text{Treatments}})=\sigma^2+\frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

We can prove this directly by apply the expected value operator. Since

$$MS_{\text{Treatments}}=\frac{SS_{\text{Treatments}}}{a-1}$$

we will initially work with the treatment sum of squares. Now

$$E(SS_{\text{Treatments}})=E\left[n\sum_{i=1}^{a}(\bar{Y}_{i.}-\bar{Y}_{..})^2\right]$$

and from the model $Y_{ij}=\mu+\tau_i+\epsilon_{ij}$ we have

$$\bar{Y}_{i.}=\mu+\tau_i+\bar{\epsilon}_{i.}$$

and

$$\bar{Y}_{..}=\mu+\bar{\epsilon}_{..}$$

since $\sum_{i=1}^{a}\tau_i=0$. Substituting for $\bar{Y}_{i.}$ and $\bar{Y}_{..}$ in the expression for $SS_{\text{Treatments}}$ yields

$$E(SS_{\text{Treatments}})=E\left[n\sum_{i=1}^{a}(\tau_i+\bar{\epsilon}_{i.}-\bar{\epsilon}_{..})^2\right]$$

$$=E\left[n\sum_{i=1}^{a}\tau_i^2+n\sum_{i=1}^{a}\bar{\epsilon}_{i.}^2+an\bar{\epsilon}_{..}^2+2n\sum_{i=1}^{a}\tau_i\bar{\epsilon}_{i.}-2n\bar{\epsilon}_{..}\sum_{i=1}^{a}\tau_i-2n\bar{\epsilon}_{..}\sum_{i=1}^{a}\bar{\epsilon}_{i.}\right]$$

However, since the ϵ_{ij}'s are independent random variables with mean zero and variance σ^2, we find that

$$E(\bar{\epsilon}_{i}^{2}.) = \frac{\sigma^2}{n}, \quad E(\bar{\epsilon}_{..}^{2}) = \frac{\sigma^2}{an}, \text{ and } E(\bar{\epsilon}_{i}.) = 0$$

Therefore

$$\begin{aligned} E(SS_{\text{Treatments}}) &= n\sum_{i=1}^{a}\tau_i^2 + a\sigma^2 - \sigma^2 \\ &= (a-1)\sigma^2 + n\sum_{i=1}^{a}\tau_i^2 \end{aligned}$$

As a result,

$$\begin{aligned} E(MS_{\text{Treatments}}) &= E\left(\frac{SS_{\text{Treatments}}}{a-1}\right) \\ &= \frac{1}{a-1}E(SS_{\text{Treatments}}) \\ &= \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1} \end{aligned}$$

Now if the null hypothesis of equal treatment means is true, each τ_i is equal to zero and

$$E\left(\frac{SS_{\text{Treatments}}}{a-1}\right) = \sigma^2$$

If the alternative hypothesis is true,

$$E\left(\frac{SS_{\text{Treatments}}}{a-1}\right) = \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

A similar approach will show that

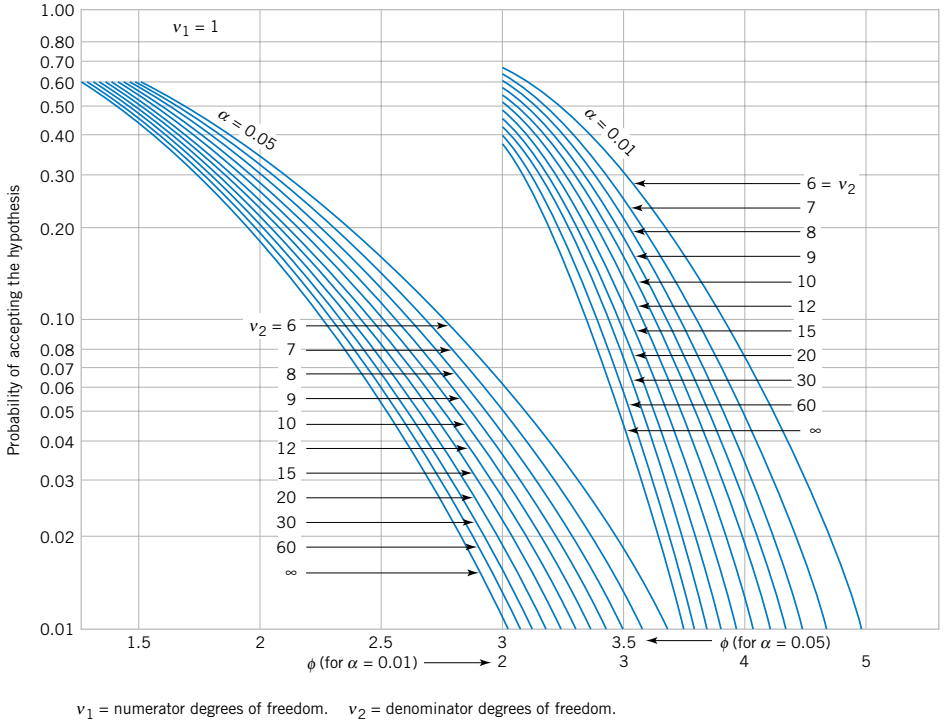
$$E(SS_E) = a(n-1)\sigma^2$$

so that

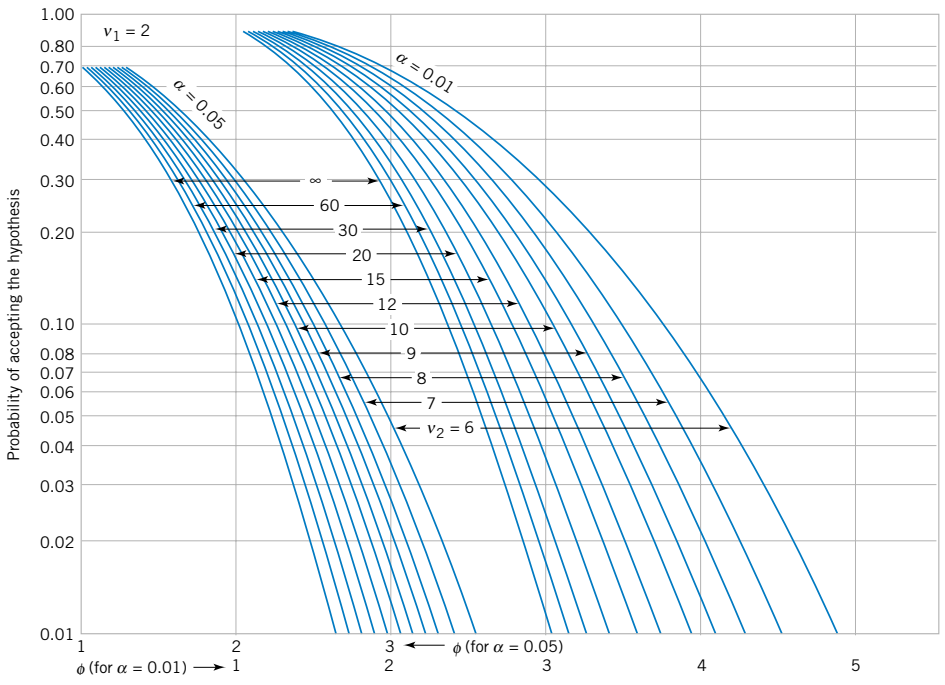
$$E(MS_E) = E\left(\frac{SS_E}{a(n-1)}\right) = \sigma^2$$

O.C. Curves

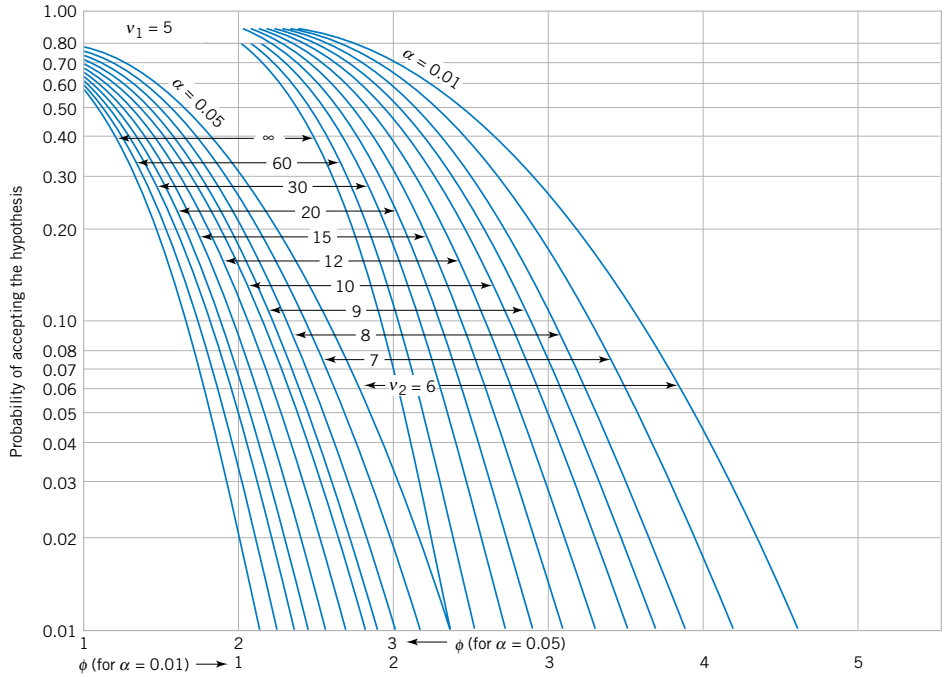
In the text, we give O.C. curves for the fixed effects ANOVA for the case of $a = 4$ and $a = 5$ treatments. A collection of additional curves for $a = 2, 3, 6, 7, 8$, and 9 are on pages 13-8 through 13-10. In using the curves, remember that v_1 = the number of numerator degrees of freedom and v_2 = the number of denominator degrees of freedom. The sample size calculation routine in Minitab will also determine sample sizes for the single-factor ANOVA.

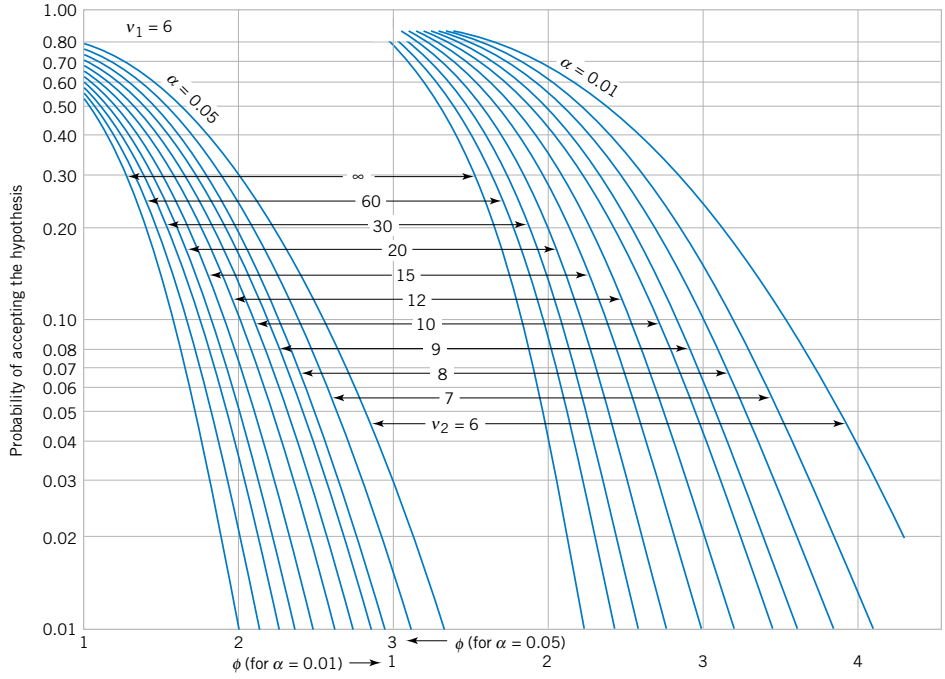


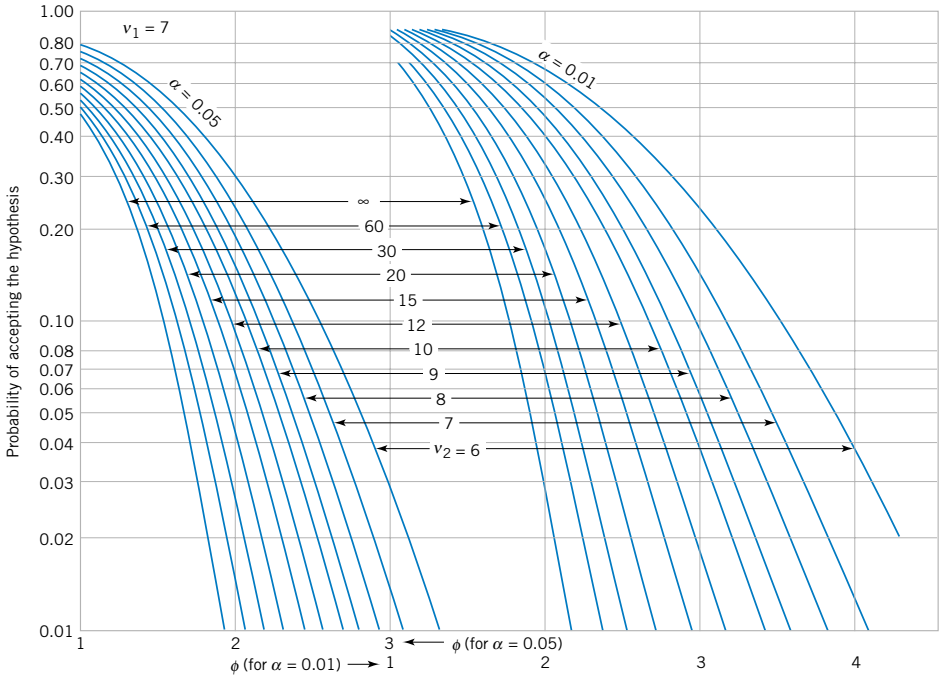
v_1 = numerator degrees of freedom. v_2 = denominator degrees of freedom.

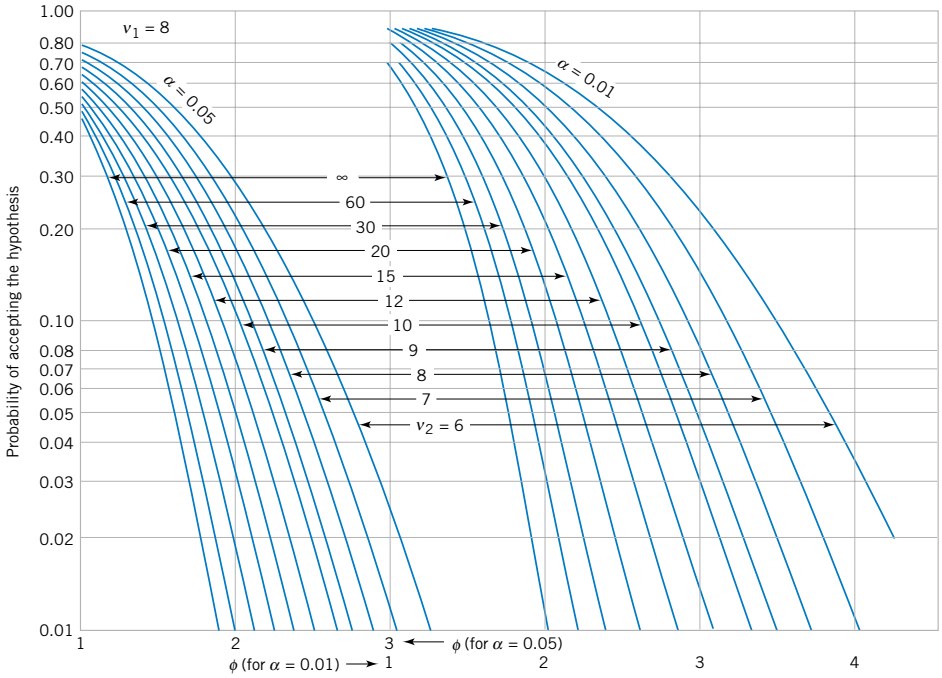


Source: These curves are adapted with permission from *Biometrika Tables for Statisticians,* Vol. 2, by E. S. Pearson and H. O. Hartley, Cambridge University Press, Cambridge, 1972.


$\nu_1 = 5$
$\alpha = 0.01$
$\alpha = 0.05$
∞
60
30
20
15
12
10
9
8
7
$\nu_2 = 6$
Probability of accepting the hypothesis
1.00
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1
2
3
ϕ (for $\alpha = 0.05$)
ϕ (for $\alpha = 0.01$)
1
2
3
4
5


$\nu_1 = 6$
$\alpha = 0.05$
$\alpha = 0.01$
∞
60
30
20
15
12
10
9
8
7
$\nu_2 = 6$
Probability of accepting the hypothesis
1.00
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1
2
3
ϕ (for $\alpha = 0.05$)
ϕ (for $\alpha = 0.01$)
1
2
3
4


$v_1 = 7$
$\alpha = 0.05$
$\alpha = 0.01$
∞
60
30
20
15
12
10
9
8
7
$v_2 = 6$
Probability of accepting the hypothesis
1.00
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1
2
3
4
ϕ (for $\alpha = 0.05$)
ϕ (for $\alpha = 0.01$)


$v_1 = 8$
$\alpha = 0.05$
$\alpha = 0.01$
∞
60
30
20
15
12
10
9
8
7
$v_2 = 6$
Probability of accepting the hypothesis
1.00
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1
2
3
4
ϕ (for $\alpha = 0.05$)
ϕ (for $\alpha = 0.01$)

13-3.3 Determining Sample Size in the Random Model (CD Only)

The power of the test for the random-effects model is

$$\begin{aligned} 1 - \beta &= P\{\text{Reject } H_0 | H_0 \text{ is false}\} \\ &= P\{F_0 > f_{\alpha,a-1,a(n-1)} | \sigma_\tau^2 > 0\} \end{aligned}$$

It can be shown that if H_1 is true ($\sigma_\tau^2 > 0$) the power can be computed using the central F distribution, with $a - 1$ and $a(n - 1)$ degrees of freedom. In fact, the ratio

$$\frac{MS_{\text{Treatments}}/(\sigma^2 + n\sigma_\tau^2)}{MS_E/\sigma^2}$$

has the F-distribution with $a - 1$ and $a(n - 1)$ degrees of freedom. Then,

$$\begin{aligned} 1 - \beta &= P\left\{\frac{MS_{\text{Treatments}}}{MS_E} > f_{\alpha,a-1,a(n-1)} \,|\, \sigma_\tau^2 > 0\right\} \\ &= P\left\{\frac{MS_{\text{Treatments}}}{MS_E(1 + n\sigma_\tau^2/\sigma^2)} > \frac{f_{\alpha,a-1,a(n-1)}}{(1 + n\sigma_\tau^2/\sigma^2)}\right\} \\ &= P\left\{F_{a-1,a(n-1)} > \frac{f_{\alpha,a-1,a(n-1)}}{(1 + n\sigma_\tau^2/\sigma^2)}\right\} \end{aligned} \tag{S13-3}$$

This probability statement may be easily evaluated using certain hand-held calculators, or it may be evaluated using tables of the F-distribution.

EXAMPLE S13-2 Consider a completely randomized design with five treatments selected at random and six observations per treatment. If $\alpha = 0.05$, what is the power of the test if $\sigma_\tau^2 = \sigma^2$?

From Equation S13-3, we have the power as

$$1 - \beta = P\left\{F_{4,25} > \frac{f_{0.05,4,25}}{[1 + 6(1)]}\right\}$$

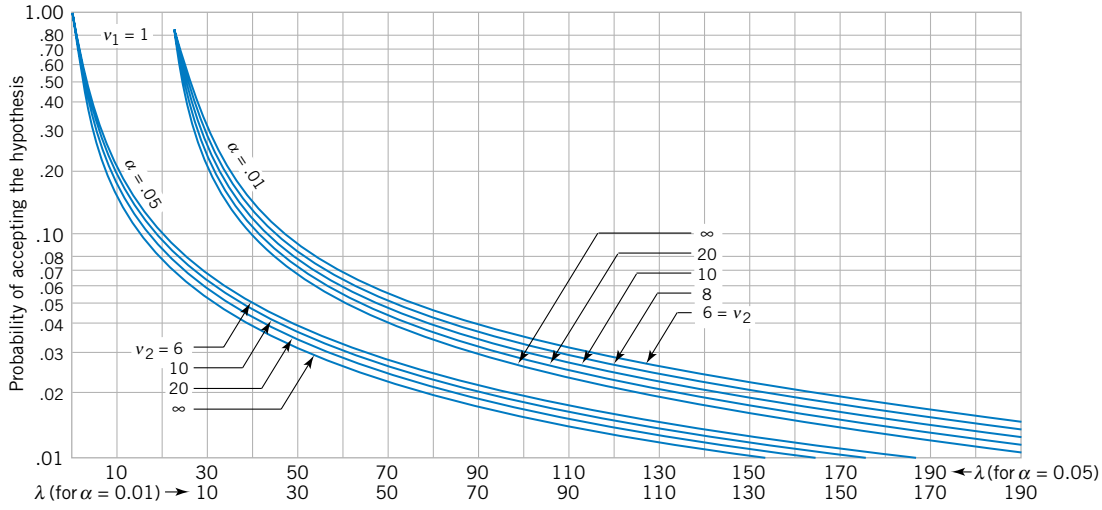
since if $\sigma_\tau^2 = \sigma^2$ the ratio $\sigma_\tau^2/\sigma^2 = 1$. Now $f_{0.05,4,25} = 2.76$, so

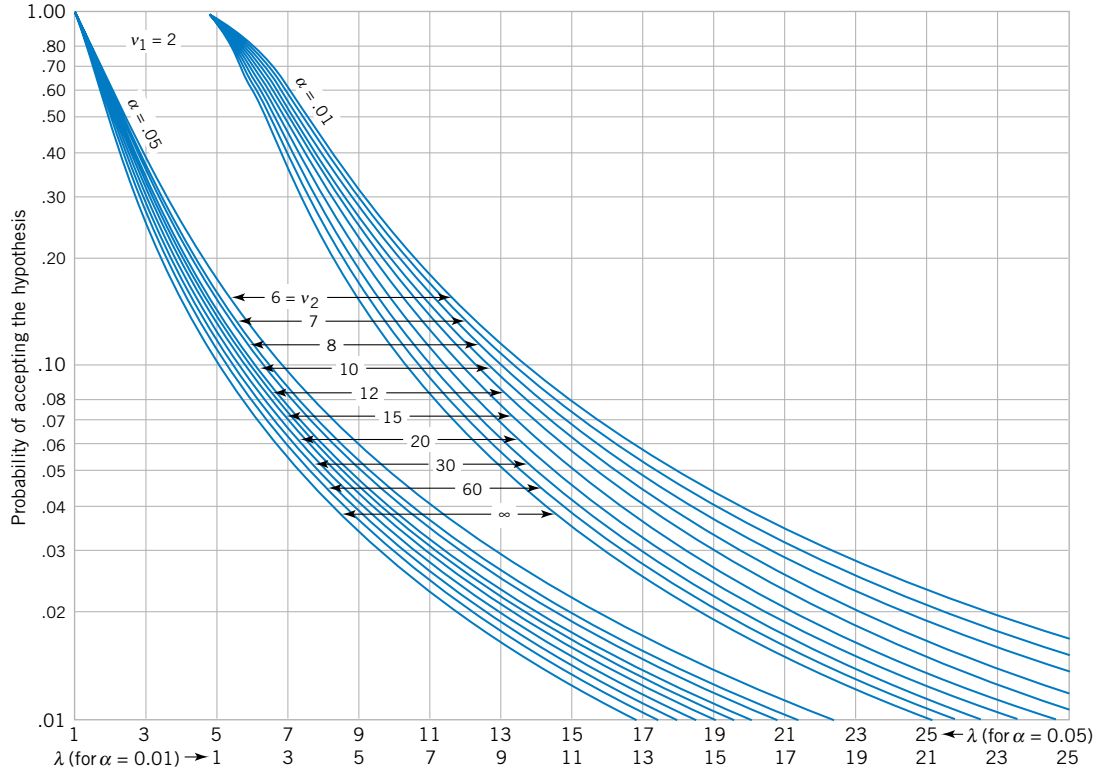
$$\begin{aligned} 1 - \beta &= F\left\{F_{4,25} > \frac{2.76}{[1 + 6(1)]}\right\} = P\left\{F_{4,25} > \frac{2.76}{7}\right\} \\ &= P\{F_{4,25} > 0.39\} = 0.81 \end{aligned}$$

This probability was evaluated using a calculator that provided F-distribution probabilities. Since the power of the test is 0.81, this implies that the null hypothesis H_0: $\sigma_\tau^2 = 0$ will be rejected with probability 0.81 in this experimental situation.

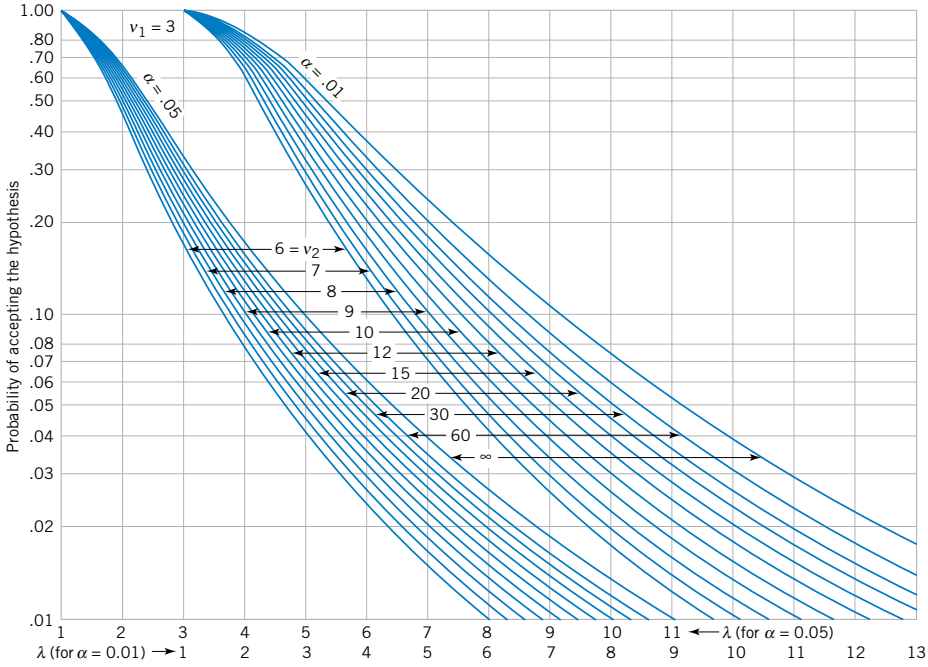
It is also possible to evaluate the power of the test using the operating characteristic curves on page 13-12 through 13-15. These curves plot the probability of the type II error β against λ, where

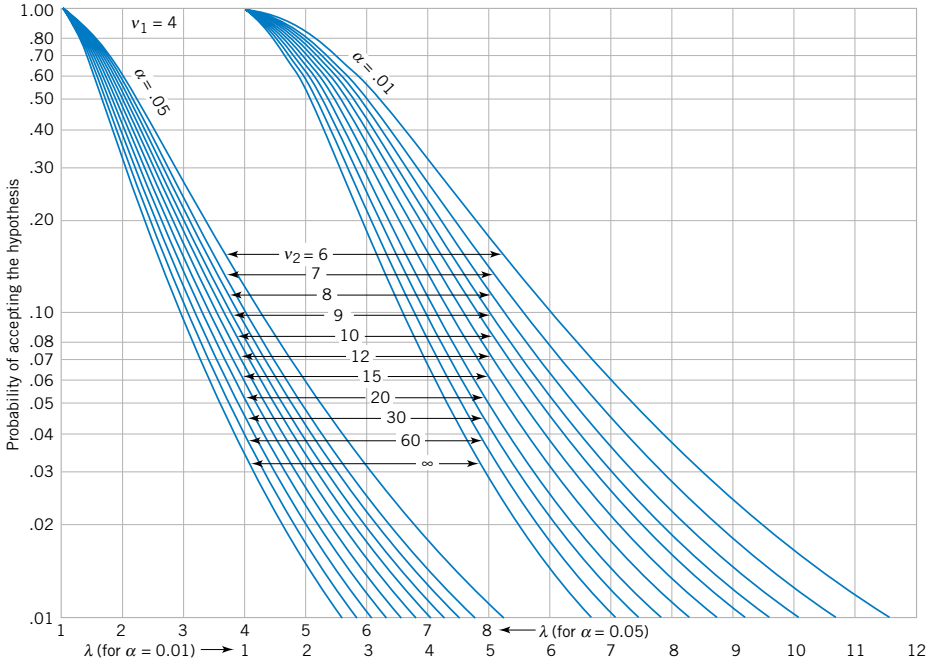
$$\lambda = \sqrt{1 + \frac{n\sigma_\tau^2}{\sigma^2}} \tag{S13-4}$$

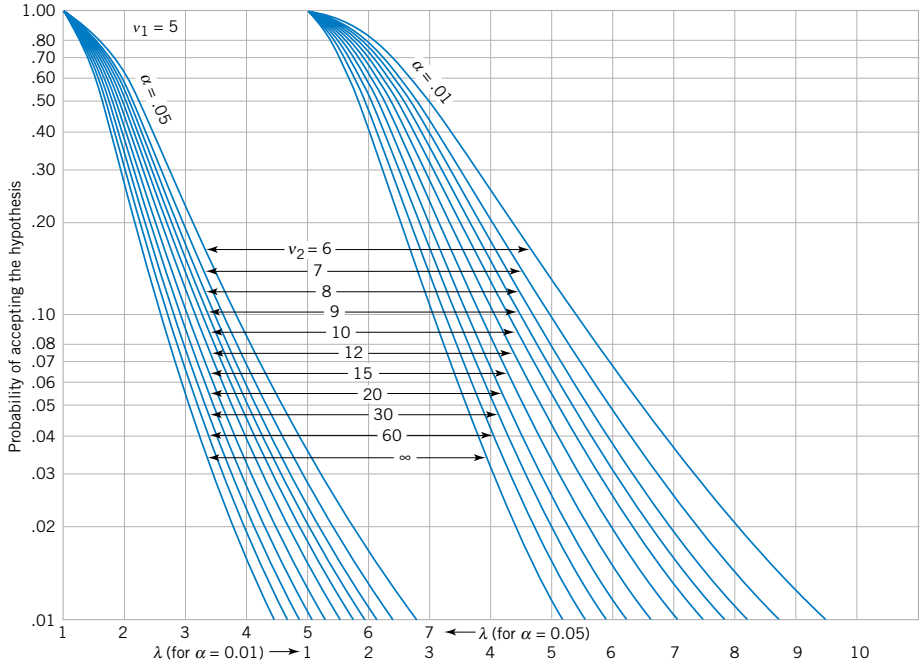


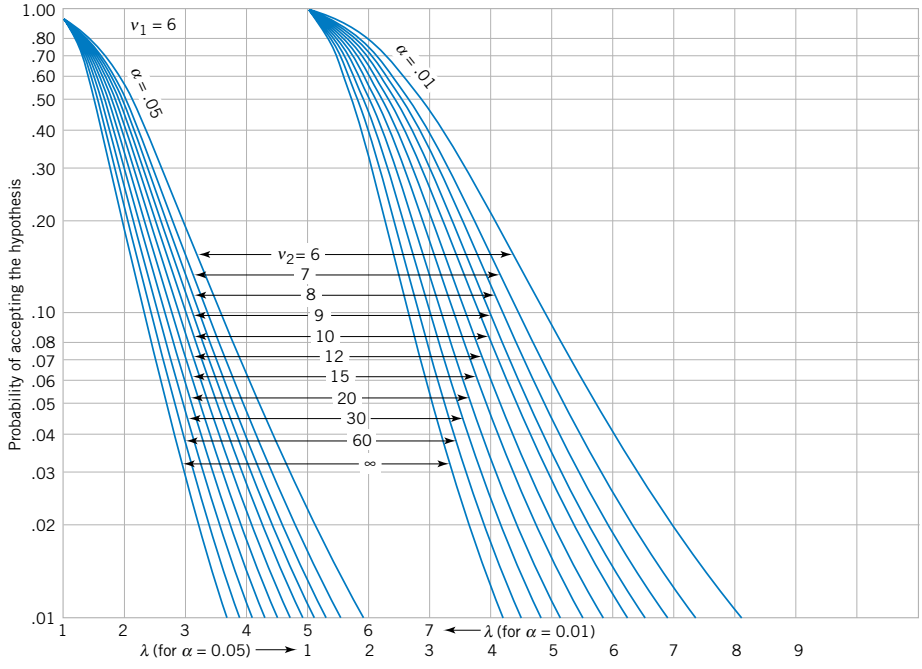


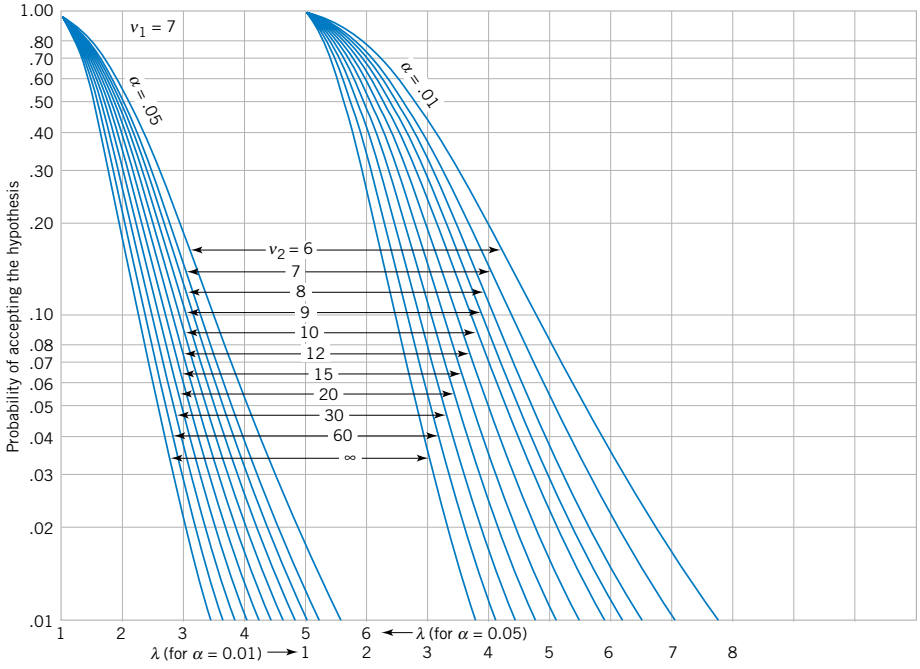
Source: Reproduced with permission from *Engineering Statistics,* 2nd edition, by A. H. Bowker and G. J. Lieberman, Prentice-Hall, Englewood Cliffs, N.J., 1972.

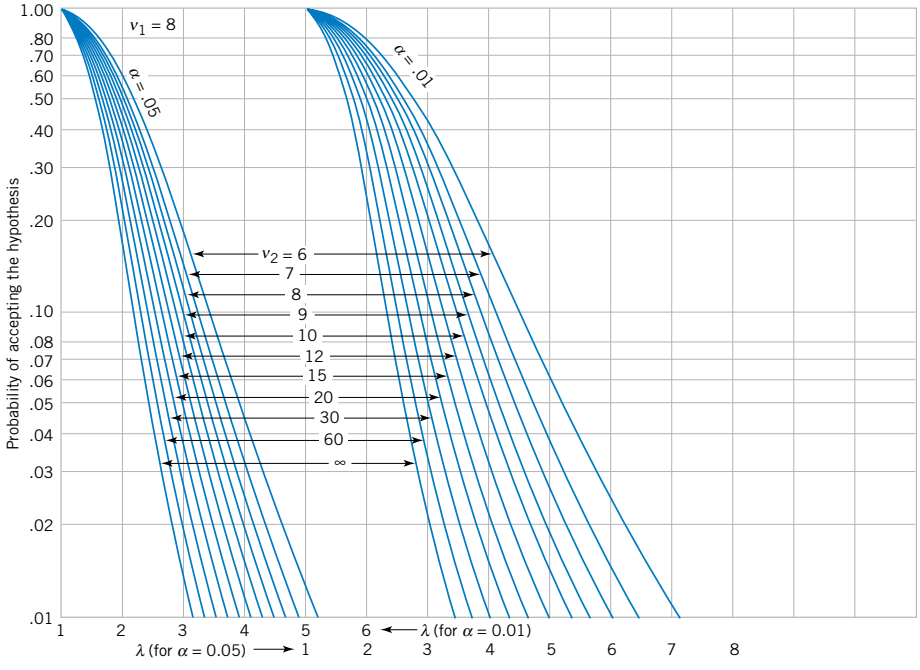

$v_1 = 3$
$\alpha = .05$
$\alpha = .01$
$6 = v_2$
7
8
9
10
12
15
20
30
60
∞
Probability of accepting the hypothesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
1 2 3 4 5 6 7 8 9 10 11 $\leftarrow \lambda$ (for $\alpha = 0.05$)
λ (for $\alpha = 0.01$) $\rightarrow$ 1 2 3 4 5 6 7 8 9 10 11 12 13


$v_1 = 4$
$\alpha = .05$
$\alpha = .01$
$v_2 = 6$
7
8
9
10
12
15
20
30
60
∞
Probability of accepting the hypothesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
1 2 3 4 5 6 7 8 $\leftarrow \lambda$ (for $\alpha = 0.05$)
λ (for $\alpha = 0.01$) $\rightarrow$ 1 2 3 4 5 6 7 8 9 10 11 12


$\nu_1 = 5$
$\alpha = .05$
$\alpha = .01$
$\nu_2 = 6$
7
8
9
10
12
15
20
30
60
∞
Probability of accepting the hypothesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
1 2 3 4 5 6 7 ← λ (for $\alpha = 0.05$)
λ (for $\alpha = 0.01$) → 1 2 3 4 5 6 7 8 9 10


$\nu_1 = 6$
$\alpha = .05$
$\alpha = .01$
$\nu_2 = 6$
7
8
9
10
12
15
20
30
60
∞
Probability of accepting the hypothesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
1 2 3 4 5 6 7 ← λ (for $\alpha = 0.01$)
λ (for $\alpha = 0.05$) → 1 2 3 4 5 6 7 8 9


$v_1 = 7$
$\alpha = .05$
$\alpha = .01$
$v_2 = 6$
7
8
9
10
12
15
20
30
60
∞
Probability of accepting the hypothesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
1 2 3 4 5 6 ← λ (for α = 0.05)
λ (for α = 0.01) → 1 2 3 4 5 6 7 8


$v_1 = 8$
$\alpha = .05$
$\alpha = .01$
$v_2 = 6$
7
8
9
10
12
15
20
30
60
∞
Probability of accepting the hypothesis
1.00
.80
.70
.60
.50
.40
.30
.20
.10
.08
.07
.06
.05
.04
.03
.02
.01
1 2 3 4 5 6 ← λ (for α = 0.01)
λ (for α = 0.05) → 1 2 3 4 5 6 7 8

In the randomized block design, replace n by b, the number of blocks. Since σ^2 is usually unknown, we may either use a prior estimate or define the value of σ_τ^2 that we are interested in detecting in terms of the ratio σ_τ^2/σ^2.

EXAMPLE S13-3 Consider the situation described in Example S13-2. Since $\sigma = 0.05$, $a = 5$, $n = 6$, and $\sigma_\tau^2 = \sigma^2$, we may find λ from Equation S13-4 as

$$\lambda = \sqrt{1 + 6(1)} = 2.646$$

From the operating characteristic curve with $v_1 = a - 1 = 4$, $v_2 = a(n-1) = 25$ degrees of freedom and $\alpha = 0.05$, we find that

$$\beta \simeq 0.20$$

Therefore, the power is approximately 0.80. This agrees with the results obtained in Example S13-2.

13-4.4 Randomized Complete Block Design with Random Factors (CD Only)

In the preceding sections, we have assumed that the treatments and blocks are fixed factors. In many randomized complete block designs, these assumptions may be too restrictive. For example, in the chemical type experiment, Example 13-5, we might like to view the fabric samples as a random sample of material to which the chemicals may be applied so that the conclusions from the experiment will extend to the entire population of material.

It turns out that, if either treatments or blocks (or both) are random effects, the F-test in the analysis of variance is still formed as

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}$$

This can be shown by using the methods presented previously to evaluate the expected mean squares. If the treatments are random, the treatment effects τ_i are considered to be normally and independently distributed random variables with mean zero and variance σ_τ^2. The null hypothesis of zero treatment effects is

$$H_0: \sigma_\tau^2 = 0$$
$$H_1: \sigma_\tau^2 > 0$$

When both treatments and blocks are random, the block effects β_j are also assumed to be normally and independently distributed random variables with mean zero and variance σ_β^2. In this case the expected values of the mean squares for treatments, blocks, and error are

$$E(MS_{\text{Treatments}}) = \sigma^2 + b\sigma_\tau^2$$
$$E(MS_{\text{Blocks}}) = \sigma^2 + a\sigma_\beta^2$$
$$E(MS_E) = \sigma^2$$

The unbiased estimates of the variance components are

$$\hat{\sigma}^2 = MS_E$$
$$\hat{\sigma}_\tau^2 = \frac{MS_{\text{Treatments}} - MS_E}{b}$$
$$\hat{\sigma}_\beta^2 = \frac{MS_{\text{Blocks}} - MS_E}{a}$$