

## Analysis of Categorical Data

1. Introduction
2. Pearson's  $\chi^2$  Test for Goodness of Fit
3. Contingency Table with One Margin Fixed  
(Test of Homogeneity)
4. Contingency Table with Neither Margin Fixed  
(Test of Independence)
5. Review Exercises

---

---

## *Are Firm Mattresses Really Best?*

Although it is commonly believed that firm mattresses can help persons with low-back pain, experimental evidence was lacking. Medical researchers<sup>1</sup> designed a carefully designed experiment. Of the patients admitted to the study, 158 were randomly selected to sleep on a new firm mattress and the others were given new medium-firm mattresses. The patients were not told about the firmness of their new mattresses although most perceived the correct firmness. The persons who installed the new beds and the research assistants who collected the data from patients were also kept unaware of the firmness of the mattresses.

This design includes meritable features: (1) the random assignment of treatments (mattress firmness) and (2) the *double blind* feature where patient and researcher are not told which treatment has been assigned.

The patients were asked to report whether or not they had low-back pain upon arising during the 90-day follow-up period.

Firmness	Low Back Pain		Total
	No	Yes	
Firm	36	122	158
Medium firm	55	95	150

The proportion with no low-back pain upon arising is  $.228 = 36/158$  for those who slept on firm mattresses and the proportion is  $.367 = 55/150$  for those who slept on medium-firm mattresses. These data provide convincing evidence that medium-firm mattresses reduce low-back pain upon arising. The  $P$ -value for the test of equality of the proportions, of persons in each population who would answer *No*, versus a two-sided alternative is .008. In less than 1 in 100 times would chance variation give an observed difference this large or larger.

<sup>1</sup>Source: F. Kovacs et al., "Effect of firmness of mattress on chronic non-specific low-back pain: randomized, double-blind, controlled, multicentre trial," *The Lancet* 362 (November 15, 2003), pp. 1599–1604.



Courtesy of Roberta Johnson.

---

---

## 1. INTRODUCTION

The expression **categorical data** refers to observations that are only classified into categories so that the data set consists of frequency counts for the categories. Such data occur abundantly in almost all fields of quantitative study, particularly in the social sciences. In a study of religious affiliations, people may be classified as Catholic, Protestant, Jewish, or other; in a survey of job compatibility, employed persons may be classified as being satisfied, neutral, or dissatisfied with their jobs; in plant breeding, the offspring of a cross-fertilization may be grouped into several genotypes; manufactured items may be sorted into such categories as “free of defects,” “slightly blemished,” and “rejects.” In all these examples, each category is defined by a qualitative trait. Categories can also be defined by specifying ranges of values on an original numerical measurement scale, such as income that is categorized high, medium, or low and rainfall that is classified heavy, moderate, or light.

The next three examples present the testing problems addressed in this chapter under the umbrella term of chi-square, or  $\chi^2$ , tests.

### Example 1 One Sample Classified in Several Categories

The offspring produced by a cross between two given types of plants can be any of the three genotypes denoted by  $A$ ,  $B$ , and  $C$ . A theoretical model of gene inheritance suggests that the offspring of types  $A$ ,  $B$ , and  $C$  should be in the ratio 1:2:1. For experimental verification, 100 plants are bred by crossing the two given types. Their genetic classifications are recorded in Table 1. Do these data contradict the genetic model?

**TABLE 1** Classification of Crossbred Plants

Genotype	$A$	$B$	$C$	Total
Observed frequency	18	55	27	100

Let us denote the population proportions or the probabilities of the genotypes  $A$ ,  $B$ , and  $C$  by  $p_A$ ,  $p_B$ , and  $p_C$ , respectively. Since the genetic model states that these probabilities are in the ratio 1:2:1, our object is to test the null hypothesis

$$H_0: p_A = \frac{1}{4} \quad p_B = \frac{2}{4} \quad p_C = \frac{1}{4}$$

Here the data consist of frequency counts of a random sample classified in three categories or cells, the null hypothesis specifies the numerical values of

the cell probabilities, and we wish to examine if the observed frequencies contradict the null hypothesis.

### Example 2 Independent Samples Classified in Several Categories

To compare the effectiveness of two diets *A* and *B*, 150 infants were included in a study. Diet *A* was given to 80 randomly selected infants and diet *B* to the other 70 infants. At a later time, the health of each infant was observed and classified into one of the three categories “excellent,” “average,” and “poor.” From the frequency counts recorded in Table 2, we wish to test the null hypothesis that there is no difference between the quality of the two diets.

**TABLE 2** Health under Two Different Diets

	Excellent	Average	Poor	Sample Size
Diet <i>A</i>	37	24	19	80
Diet <i>B</i>	17	33	20	70
Total	54	57	39	150

The two rows of Table 2 have resulted from independent samples. For a descriptive summary of these data, it is proper to compute the relative frequencies for each row. These are given in Table 2(a).

**TABLE 2 (a)** Relative Frequencies (from Table 2)

	Excellent	Average	Poor	Total
Diet <i>A</i>	.46	.30	.24	1
Diet <i>B</i>	.24	.47	.29	1

The (unknown) population proportions or probabilities are entered in Table 2(b). They allow us to describe the null hypothesis more clearly. The null hypothesis of “no difference” is equivalent to the statement that, for each response category, the probability is the same for diet *A* and diet *B*. Consequently, we formulate

$$H_0: p_{A1} = p_{B1} \quad p_{A2} = p_{B2} \quad p_{A3} = p_{B3}$$

Note that although  $H_0$  specifies a structure for the cell probabilities, it does not give the numerical value of the common probability in each column.

**TABLE 2(b)** Population Proportions or Probabilities

	Excellent	Average	Poor	Total
Diet A	$p_{A1}$	$p_{A2}$	$p_{A3}$	1
Diet B	$p_{B1}$	$p_{B2}$	$p_{B3}$	1

### Example 3 One Sample Simultaneously Classified According to Two Characteristics

Improperly disclosed personal information has become a major source leading to credit fraud. In a survey conducted to help investigate the pervasiveness of the problem, 395 respondents reported receiving notification that their personal data had been compromised. Each of these cases were then classified according to type of institution sending the notification and also whether or not harm was caused to the person's credit. Harm to credit primarily includes instances of charging merchandise or taking money from bank accounts. Attention was restricted to institutions categorized as

Government (federal, state and local)

Financial (bank, investment organization, credit firm, or insurance)

Other Commercial (Internet, retail, or communications company).

The cell frequencies for the two-way classification<sup>2</sup> are given in Table 3.

Do the data indicate that the pattern of cases of harmed credit, caused by the release of personal data, differs among the three types of organizations?

**TABLE 3** Type of Organization and Fraudulent Use of Personal Information

	Government	Financial	Other Commercial	Total
Harm	30	24	21	75
No Harm	182	104	34	320
Total	212	128	55	395

Unlike Example 2, here we have a single random sample, but each sampled individual elicits two types of response: source of notification and whether or not harm was done to credit. In the present context, the null hypothesis of "no difference" amounts to saying the outcome on credit and type of institution giving notification are independent. In other words, whether or not harm was done is unrelated to, or independent of, the type of institution that improperly disclosed personal information. A formal specification of this null hypothesis, in terms of the cell probabilities, is deferred until Section 4.

<sup>2</sup>The resulting proportions are close to those reported in a 2006 survey by Harris Interactive.

Frequency count data that arise from a classification of the sample observations according to two or more characteristics are called **cross-tabulated data** or a **contingency table**. If only two characteristics are observed and the contingency table has  $r$  rows and  $c$  columns, it is designated as an  $r \times c$  table. Thus, Tables 2 and 3 are both  $2 \times 3$  contingency tables.

Although Tables 2 and 3 have the same appearance, there is a fundamental difference in regard to the method of sampling. The row totals 80 and 70 in Table 2 are the predetermined sample sizes; these are not outcomes of random sampling, as are the column totals. By contrast, both sets of marginal totals in Table 3 are outcomes of random sampling—none were fixed beforehand. To draw the distinction, one often refers to Table 2 as a  $2 \times 3$  **contingency table** with fixed row totals. In Sections 3 and 4, we will see that the formulation of the null hypothesis is different for the two situations.

## 2. PEARSON'S $\chi^2$ TEST FOR GOODNESS OF FIT

We first consider the type of problem illustrated in Example 1, where the data consist of frequency counts observed from a random sample and the null hypothesis specifies the unknown cell probabilities. Our primary goal is to test if the model given by the null hypothesis fits the data, and this is appropriately called **testing for goodness of fit**.

For general discussion, suppose a random sample of size  $n$  is classified into  $k$  categories or cells labeled  $1, 2, \dots, k$  and let  $n_1, n_2, \dots, n_k$  denote the respective cell frequencies. If we denote the cell probabilities by  $p_1, p_2, \dots, p_k$ , a null hypothesis that completely specifies the cell probabilities is of the form

$$H_0: p_1 = p_{10}, \dots, p_k = p_{k0}$$

where  $p_{10}, \dots, p_{k0}$  are given numerical values that satisfy  $p_{10} + \dots + p_{k0} = 1$ .

From Chapter 5 recall that if the probability of an event is  $p$ , then the expected number of occurrences of the event in  $n$  trials is  $np$ . Therefore, once the cell probabilities are specified, the expected cell frequencies can be readily computed by **multiplying** these probabilities by the sample size  $n$ . A goodness of fit test attempts to determine if a conspicuous discrepancy exists between the observed cell frequencies and those expected under  $H_0$ . (See Table 4.)

**TABLE 4** The Basis of a Goodness of Fit Test

Cells	1	2	...	$k$	Total
Observed frequency $O$	$n_1$	$n_2$	...	$n_k$	$n$
Probability under $H_0$	$p_{10}$	$p_{20}$	...	$p_{k0}$	1
Expected frequency $E$ under $H_0$	$np_{10}$	$np_{20}$	...	$np_{k0}$	$n$

A useful measure for the overall discrepancy between the observed and expected frequencies is given by the **chi-square** or  **$\chi^2$  statistic**.

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_{i0})^2}{np_{i0}} = \sum_{\text{cells}} \frac{(O - E)^2}{E}$$

where  $O$  and  $E$  symbolize an observed frequency and the corresponding expected frequency. The discrepancy in each cell is measured by the squared difference between the observed and the expected frequencies divided by the expected frequency. The  $\chi^2$  measure is the sum of these quantities for all cells.

The  $\chi^2$  statistic was originally proposed by Karl Pearson (1857–1936), who found the distribution for large  $n$  to be approximately a  $\chi^2$  distribution with d.f. =  $k - 1$ . Due to this distribution, the statistic is denoted by  $\chi^2$  and called **Pearson's  $\chi^2$  statistic for goodness of fit**. Because a large value of the overall discrepancy indicates a disagreement between the data and the null hypothesis, the upper tail of the  $\chi^2$  distribution constitutes the rejection region.

### Pearson's $\chi^2$ Test for Goodness of Fit (Based on Large $n$ )

Null hypothesis

$$H_0: p_1 = p_{10}, \dots, p_k = p_{k0}$$

Test statistic

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_{i0})^2}{np_{i0}} = \sum_{\text{cells}} \frac{(O - E)^2}{E}$$

Rejection region

$$\chi^2 \geq \chi_{\alpha}^2$$

where  $\chi_{\alpha}^2$  is the upper  $\alpha$  point of the  $\chi^2$  distribution with

$$\text{d.f.} = k - 1 = (\text{Number of cells}) - 1$$

It should be remembered that Pearson's  $\chi^2$  test is an approximate test that is valid only for large samples. As a rule of thumb,  $n$  should be large enough so that the expected frequency of each cell is at least 5.

#### Example 4 $\chi^2$ Goodness of Fit for a Genetic Model

Referring to Example 1, test the goodness of fit of the genetic model to the data in Table 1. Take  $\alpha = .05$ .

**SOLUTION** Following the structure of Table 4, the computations for the  $\chi^2$  statistic are exhibited in Table 5 where the last line gives the calculation

$$\begin{aligned} \sum_{\text{cells}} \frac{(O - E)^2}{E} &= \frac{(18 - 25)^2}{25} + \frac{(55 - 50)^2}{50} + \frac{(27 - 25)^2}{25} \\ &= 1.96 \quad + \quad .50 \quad + \quad .16 \quad = 2.62 \end{aligned}$$

**TABLE 5** The  $\chi^2$  Goodness of Fit Test for the Data in Table 1

Cell	A	B	C	Total
Observed frequency $O$	18	55	27	100
Probability under $H_0$	.25	.50	.25	1.0
Expected frequency $E$	25	50	25	100
$\frac{(O - E)^2}{E}$	1.96	.50	.16	2.62 = $\chi^2$ d.f. = 2

We use the  $\chi^2$  statistic with rejection region  $R: \chi^2 \geq 5.99$  since  $\chi_{.05}^2 = 5.99$  with d.f. = 2 (Appendix B, Table 5). Because the observed  $\chi^2 = 2.62$  is smaller than this value, the null hypothesis is not rejected at  $\alpha = .05$ . We conclude that the data in Table 1 do not contradict the genetic model.

The  $\chi^2$  statistic measures the overall discrepancy between the observed frequencies and those expected under a given null hypothesis. Example 4 demonstrates its application when the frequency counts arise from a single random sample and the categories refer to only one characteristic—namely, the genotype of the offspring. Basically, the same principle extends to testing hypotheses with more complex types of categorical data such as the contingency tables illustrated in Examples 2 and 3. In preparation for these developments, we state two fundamental properties of the  $\chi^2$  statistic:

### Properties of Pearson's $\chi^2$ Statistic

1. **Additivity:** If  $\chi^2$  statistics are computed from independent samples, then their sum is also a  $\chi^2$  statistic whose d.f. equals the sum of the d.f.'s of the components.
2. **Loss of d.f. due to estimation of parameters:** If  $H_0$  does not completely specify the cell probabilities, then some parameters have to be estimated in order to obtain the expected cell frequencies. In that case, the d.f. of  $\chi^2$  is reduced by the number of parameters estimated.

$$\text{d.f. of } \chi^2 = (\text{No. of cells}) - 1 - (\text{No. of parameters estimated})$$



**Exercises**

13.1 Given below are the frequencies observed from 320 tosses of a die. Do these data cast doubt on the fairness of the die?

Face No.	1	2	3	4	5	6	Total
Frequency	39	63	56	67	57	38	320

13.2 Recorded here is the frequency distribution of the blood types of 100 persons who have volunteered to donate blood at a plasma center.

Blood type	O	A	B	AB	Total
Frequency	40	44	10	6	100

Test the goodness of fit of the model, which assumes that the four blood types are equally likely in the population of plasma donors. Use  $\alpha = .05$ .

13.3 Referring to the data in Exercise 13.2, test the null hypothesis that the probability of the blood types O, A, B, and AB is in the ratios 4:4:1:1. Use  $\alpha = .05$ .

13.4 A market researcher wishes to assess consumers' preference among three different colors available on name-brand dishwashers. The following frequencies were observed from a random sample of 150 recent sales.

Color	White	Stainless	Black	Total
Frequency	63	56	31	150

Test the null hypothesis, at  $\alpha = .05$ , that all three colors are equally popular.

13.5 A shipment of assorted nuts is labeled as having 45% walnuts, 20% hazelnuts, 20% almonds, and 15% pistachios. By randomly picking several scoops of nuts from this shipment, an inspector finds the counts shown at the top of next column. Could these findings be a strong basis for an accusation of mislabeling? Test at  $\alpha = .025$ .

	Counts
Walnuts	95
Hazelnuts	70
Almonds	33
Pistachios	42
Total	240

13.6 Cross-fertilizing a pure strain of red flowers with a pure strain of white flowers produces pink hybrids that have one gene of each type. Crossing these hybrids can lead to any one of four possible gene pairs. Under Mendel's theory, these four are equally likely, so

$$P(\text{white}) = \frac{1}{4} \quad P(\text{pink}) = \frac{1}{2} \quad P(\text{red}) = \frac{1}{4}$$

An experiment carried out by Correns, one of Mendel's followers, resulted in the frequencies 141, 291, and 132 for the white, pink, and red flowers, respectively. (Source: W. Johannsen, 1909, *Elements of the Precise Theory of Heredity*, G. Fischer, Jena.)

Do these observations appear to contradict the probabilities suggested by Mendel's theory?

13.7 According to the records of the National Safety Council, accidental deaths in the United States during 2005 had the following distribution according to the principal types of accidents.

Motor Vehicle	Poison	Falls	Choking	Drowning	Other
40.5%	18.5%	15.7%	4.1%	3.2%	18.0%

In a particular geographical region, the accidental deaths, classified according to the principal types of accidents, yielded the following frequency distribution.

Motor Vehicle	Poison	Falls	Choking	Drowning	Other	Total
356	207	125	33	26	161	908

Do these data show a significantly different distribution of accidental deaths as compared to that for the entire United States? Test at  $\alpha = .05$ .

- 13.8 The following table, based on government data, shows the frequency distribution of births by day of the week for all registered births in the United States in a recent year. Test the null hypothesis, at  $\alpha = .01$ , that all seven days of the week are equally likely for childbirth.

Number of Births (in 10,000) by Day of the Week, United States

	Number of births
Mon.	62.81
Tues.	69.66
Wed.	70.11
Thurs.	69.91
Fri.	68.39
Sat.	45.47
Sun.	40.21
All Days	426.56

Source: J. Martin et al, "Births: Final Data for 2006," *National Vital Statistics Reports* 57, No. 7 (2009), Table 17.

- 13.9 Observations of 80 litters, each containing 3 rabbits, reveal the following frequency distribution of the number of male rabbits per litter.

Number of males in litter	0	1	2	3	Total
Number of litters	19	32	22	7	80

### 3. CONTINGENCY TABLE WITH ONE MARGIN FIXED (TEST OF HOMOGENEITY)

From each population, we draw a random sample of a predetermined sample size and classify each response in categories. These data form a two-way contingency table where one classification refers to the populations and the other to the response under study. Our objective is to test whether the populations are alike, or **homogeneous**, with respect to cell probabilities. To do so, we will determine if the observed proportions in each response category are nearly the same for all populations.

Let us pursue our development of the  $\chi^2$  test of homogeneity with the data of Table 2.

#### Example 5 Developing a $\chi^2$ Test to Compare Two Diets

Referring to Example 2, test the null hypothesis that there is no difference between the quality of the two diets.

Under the model of Bernoulli trials for the sex of rabbits, the probability distribution of the number of males per litter should be binomial with 3 trials and  $p =$  probability of a male birth. From these data, the parameter  $p$  is estimated as

$$\begin{aligned}\hat{p} &= \frac{\text{Total number of males in 80 litters}}{\text{Total number of rabbits in 80 litters}} \\ &= \frac{97}{240} \approx .4\end{aligned}$$

- (a) Using the binomial table for three trials and  $p = .4$ , determine the cell probabilities.  
 (b) Perform the  $\chi^2$  test for goodness of fit. (In determining the d.f., note that one parameter has been estimated from the data.)
- \*13.10 *An alternative expression for Pearson's  $\chi^2$ .* By expanding the square on the right-hand side of

$$\chi^2 = \sum_{\text{cells}} \frac{(n_i - np_{i0})^2}{np_{i0}}$$

show that the  $\chi^2$  statistic can also be expressed as

$$\begin{aligned}\chi^2 &= \sum_{\text{cells}} \frac{n_i^2}{np_{i0}} - n \quad \text{that is,} \\ \chi^2 &= \sum_{\text{cells}} \frac{O^2}{E} - n\end{aligned}$$

**SOLUTION** For ease of reference, the data in Table 2 are reproduced in Table 6. Here the populations correspond to the two diets and the response is recorded in three categories. The row totals 80 and 70 are the fixed sample sizes.

**TABLE 6** A  $2 \times 3$  Contingency Table with Fixed Row Totals

	Excellent	Average	Poor	Total
Diet A	37	24	19	80
Diet B	17	33	20	70
Total	54	57	39	150

We have already formulated the null hypothesis of “homogeneity” or “no difference between the diets” as [see Table 2(b)]

$$H_0: p_{A1} = p_{B1} \quad p_{A2} = p_{B2} \quad p_{A3} = p_{B3}$$

If we denote these common probabilities under  $H_0$  by  $p_1$ ,  $p_2$ , and  $p_3$ , respectively, the expected cell frequencies in each row would be obtained by multiplying these probabilities by the sample size. In particular, the expected frequencies in the first row are  $80 p_1$ ,  $80 p_2$ , and  $80 p_3$ , and those in the second row are  $70 p_1$ ,  $70 p_2$ , and  $70 p_3$ . However, the  $p_i$ 's are not specified by  $H_0$ . Therefore, we have to estimate these parameters in order to obtain the numerical values of the expected frequencies.

The column totals 54, 57, and 39 in Table 6 are the frequency counts of the three response categories in the combined sample of size 150. Under  $H_0$ , the estimated probabilities are

$$\hat{p}_1 = \frac{54}{150} \quad \hat{p}_2 = \frac{57}{150} \quad \hat{p}_3 = \frac{39}{150}$$

We use these estimates to calculate the expected frequencies in the first row as

$$80 \times \frac{54}{150} = \frac{80 \times 54}{150} = 28.8 \quad \frac{80 \times 57}{150} = 30.4 \quad \frac{80 \times 39}{150} = 20.8$$

and similarly for the second row. Referring to Table 6, notice the interesting pattern in these calculations:

$$\text{Expected cell frequency} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

Table 7(a) presents the observed frequencies  $O$  along with the expected frequencies  $E$ . The latter are given in parentheses. Table 7(b) computes the discrepancy measure  $(O - E)^2 / E$  for the individual cells. Adding these over all the cells, we obtain the value of the  $\chi^2$  statistic.

$$\begin{aligned} \chi^2 &= 2.335 + 1.347 + .156 + \\ &\quad 2.668 + 1.540 + .178 \quad = 8.224 \end{aligned}$$

**TABLE 7(a)** The Observed and Expected Frequencies of the Data in Table 6

	Excellent	Average	Poor
Diet A	37 (28.8)	24 (30.4)	19 (20.8)
Diet B	17 (25.2)	33 (26.6)	20 (18.2)

**TABLE 7(b)** The Values of  $(O - E)^2 / E$ 

	Excellent	Average	Poor	
Diet A	2.335	1.347	.156	
Diet B	2.668	1.540	.178	
				8.224 = $\chi^2$

In order to determine the degrees of freedom, we employ the properties of the  $\chi^2$  statistic stated in Section 2. Our  $\chi^2$  statistic has been computed from two independent samples; each contributes  $3 - 1 = 2$  d.f. because there are three categories. The added d.f. =  $2 + 2 = 4$  must now be reduced by the number of parameters we have estimated. Since  $p_1$ ,  $p_2$ , and  $p_3$  satisfy the relation  $p_1 + p_2 + p_3 = 1$ , there are really two undetermined parameters among them. Therefore, our  $\chi^2$  statistic has d.f. =  $4 - 2 = 2$ .

With d.f. = 2, the tabulated upper 5% point of  $\chi^2$  is 5.99 (Appendix B, Table 5). Since the observed  $\chi^2 = 8.224$  is larger, the null hypothesis is rejected at  $\alpha = .05$ . It would also be rejected at  $\alpha = .025$ . Therefore, a significant difference between the quality of the two diets is indicated by the data.

Having obtained a significant  $\chi^2$ , we should now examine Tables 7(a) and 7(b) and try to locate the source of the significance. We find that large contributions to  $\chi^2$  come from the "excellent" category, where the relative frequency is  $37/80$ , or 46%, for diet A and  $17/70$ , or 24%, for diet B. These data indicate that A is superior.

Motivated by Example 5, we are now ready to describe the  $\chi^2$  test procedure for an  $r \times c$  contingency table that has independent samples from  $r$  populations that are classified in  $c$  response categories. As we have seen before, the expected frequency of a cell is given by (row total  $\times$  column total)/grand

total. With regard to the d.f. of the  $\chi^2$  for an  $r \times c$  table, we note that each of the  $r$  rows contributes  $c - 1$  d.f.'s so the total contribution is  $r(c - 1)$ . Since  $c - 1$  number of parameters have to be estimated,

$$\begin{aligned} \text{d.f. of } \chi^2 &= r(c - 1) - (c - 1) \\ &= (r - 1)(c - 1) \\ &= (\text{No. of rows} - 1) \times (\text{No. of columns} - 1) \end{aligned}$$

### The $\chi^2$ Test of Homogeneity in a Contingency Table

#### Null hypothesis

In each response category, the probabilities are equal for all the populations.

#### Test statistic

$$\chi^2 = \sum_{\text{cells}} \frac{(O - E)^2}{E} \quad \left\{ \begin{array}{l} O = \text{Observed cell frequency} \\ E = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}} \end{array} \right.$$

$$\text{d.f.} = (\text{No. of rows} - 1) \times (\text{No. of columns} - 1)$$

#### Rejection region

$$\chi^2 \geq \chi^2_{\alpha}$$

### Example 6 Conducting a $\chi^2$ Test of Homogeneity

A survey is undertaken to determine the incidence of alcoholism in different professional groups. Random samples of the clergy, educators, executives, and merchants are interviewed, and the observed frequency counts are given in Table 8.

**TABLE 8** Contingency Table of Alcoholism versus Profession

	Alcoholic	Nonalcoholic	Sample Size
Clergy	32 (58.25)	268 (241.75)	300
Educators	51 (48.54)	199 (201.46)	250
Executives	67 (58.25)	233 (241.75)	300
Merchants	83 (67.96)	267 (282.04)	350
Total	233	967	1200

Construct a test to determine if the incidence rate of alcoholism appears to be the same in all four groups.

**SOLUTION** Let us denote the proportions of alcoholics in the populations of the clergy, educators, executives, and merchants by  $p_1$ ,  $p_2$ ,  $p_3$ , and  $p_4$ , respectively. Based on independent random samples from four binomial populations, we want to test the null hypothesis

$$H_0: p_1 = p_2 = p_3 = p_4$$

The expected cell frequencies, shown in parentheses in Table 8, are computed by multiplying the row and column totals and dividing the results by 1200. The  $\chi^2$  statistic is computed in Table 9. With d.f. = 3, the tabulated upper 5% point of  $\chi^2$  is 7.81 so that the null hypothesis is rejected at  $\alpha = .05$ . It would be rejected also at  $\alpha = .01$ , so the  $P$ -value is less than .01.

**TABLE 9** The Values of  $(O - E)^2/E$  for the Data in Table 8

	Alcoholic	Nonalcoholic	
Clergy	11.83	2.85	
Educators	.12	.03	
Executives	1.31	.32	
Merchants	3.33	.80	
			20.59 = $\chi^2$
			d.f. of $\chi^2 = (4 - 1)(2 - 1) = 3$

Examining Table 9, we notice that a large contribution to the  $\chi^2$  statistic has come from the first row. This is because the relative frequency of alcoholics among the clergy is quite low in comparison to the others, as one can see from Table 8.

### Example 7 The $\chi^2$ Test for $2 \times 2$ Contingency Table

To determine the possible effect of a chemical treatment on the rate of seed germination, 100 chemically treated seeds and 150 untreated seeds are sown. The numbers of seeds that germinate are recorded in Table 10. Do the data provide strong evidence that the rate of germination is different for the treated and untreated seeds?

**TABLE 10** Germination of Seeds

	Germinated	Not Germinated	Total
Treated	84(86.40)	16(13.60)	100
Untreated	132(129.60)	18(20.40)	150
Total	216	34	250

**SOLUTION** Letting  $p_1$  and  $p_2$  denote the probabilities of germination for the chemically treated seeds and the untreated seeds, respectively, we wish to test the null hypothesis  $H_0: p_1 = p_2$  versus  $H_1: p_1 \neq p_2$ . For the  $\chi^2$  test, we calculate the expected frequencies in the usual way. These are given in parentheses in Table 10. The computed value of  $\chi^2$  is

$$\begin{aligned}\chi^2 &= .067 + .424 + .044 + .282 \\ &= .817\end{aligned}$$

$$\text{d.f.} = (2 - 1)(2 - 1) = 1$$

The tabulated 5% value of  $\chi^2$  with d.f. = 1 is 3.84. Because the observed  $\chi^2 = .817$  is smaller, the null hypothesis is not rejected at  $\alpha = .05$ . The rate of germination is not significantly different between the treated and untreated seeds.

### ANOTHER METHOD OF ANALYZING A $2 \times 2$ CONTINGENCY TABLE

In light of Example 7, we note that a  $2 \times 2$  contingency table, with one margin fixed, is essentially a display of independent random samples from two dichotomous (i.e., two-category) populations. This structure is shown in Table 11, where we have labeled the two categories “success” and “failure.” Here  $X$  and  $Y$  denote the numbers of successes in independent random samples of sizes  $n_1$  and  $n_2$  taken from population 1 and population 2, respectively.

**TABLE 11** Independent Samples  
from Two Dichotomous Populations

	No. of Successes	No. of Failures	Sample Size
Population 1	$X$	$n_1 - X$	$n_1$
Population 2	$Y$	$n_2 - Y$	$n_2$

If we let  $p_1$  and  $p_2$  denote the probabilities of success for populations 1 and 2, respectively, our objective is to test the null hypothesis  $H_0: p_1 = p_2$ . The sample proportions

$$\hat{p}_1 = \frac{X}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{Y}{n_2}$$

provide estimates of  $p_1$  and  $p_2$ . When the sample sizes are large, a test of  $H_0: p_1 = p_2$  can be based on (see Section 6 of Chapter 10)

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{where } \hat{p} = \frac{X + Y}{n_1 + n_2}$$

Test statistic

The level  $\alpha$  rejection region is  $|Z| \geq z_{\alpha/2}$ ,  $Z \leq -z_{\alpha}$ , or  $Z \geq z_{\alpha}$  according to whether the alternative hypothesis is  $p_1 \neq p_2$ ,  $p_1 < p_2$ , or  $p_1 > p_2$ . Here  $z_{\alpha}$  denotes the upper  $\alpha$  point of the  $N(0, 1)$  distribution.

Although the test statistics  $Z$  and

$$\chi^2 = \sum_{\text{cells}} \frac{(O - E)^2}{E}$$

appear to have quite different forms, there is an exact relation between them—namely,

$$Z^2 = \chi^2 \quad (\text{for a } 2 \times 2 \text{ contingency table})$$

Also,  $z_{\alpha/2}^2$  is the same as the upper  $\alpha$  point of  $\chi^2$  with d.f. = 1. For instance, with  $\alpha = .05$ ,  $z_{.025}^2 = (1.96)^2 = 3.8416$ , which is also the upper 5% point of  $\chi^2$  with d.f. = 1 (see Appendix B, Table 5). Thus, the two test procedures are equivalent when the alternative hypothesis is two-sided. However, if the alternative hypothesis is one-sided, such as  $H_1 : p_1 > p_2$ , only the  $Z$  test is appropriate.

### Example 8 Conducting a $Z$ test to Compare Two Proportions

Use the  $Z$  test with the data of Example 7. Also, determine a 95% confidence interval for  $p_1 - p_2$ .

**SOLUTION** We calculate

$$\hat{p}_1 = \frac{84}{100} = .84 \quad \hat{p}_2 = \frac{132}{150} = .88$$

$$\text{Pooled estimate } \hat{p} = \frac{84 + 132}{100 + 150} = .864$$

$$\begin{aligned} Z &= \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= \frac{-.04}{\sqrt{.864 \times .136} \sqrt{\frac{1}{100} + \frac{1}{150}}} = -.904 \end{aligned}$$

Because the observed  $|Z|$  is smaller than  $z_{.025} = 1.96$ , the null hypothesis is not rejected at  $\alpha = .05$ . Note that  $Z^2 = (-.904)^2 = .817$  agrees with the result  $\chi^2 = .817$  found in Example 7.



For the confidence interval, we calculate

$$\hat{p}_1 - \hat{p}_2 = -.04$$

$$\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} = \sqrt{\frac{.84 \times .16}{100} + \frac{.88 \times .12}{150}}$$

$$= .045$$

A 95% confidence interval for  $p_1 - p_2$  is

$$-.04 \pm 1.96 \times .045 = -.04 \pm .09 \quad \text{or} \quad (-.13, .05)$$

### Exercises

13.11 Among a sample of 800 adult males, 414 said they usually open all of their mail. Among 900 adult females, 532 said they usually open all of their mail.<sup>3</sup>

- (a) Construct a two-way table based on these frequencies.
- (b) Formulate the null hypothesis.
- (c) Conduct a  $\chi^2$  test of your null hypothesis. Use  $\alpha = .05$ .
- (d) Comment on the form of any departure from the null hypothesis.

13.12 Chapter 3, Table 1, gives the counts of underclassmen and for upperclassmen according to hours worked.

	No Job	Work 10 Hours or Less	Work More Than 10 Hours	Total
underclassmen	132	28	20	180
Upperclassmen	124	44	52	220
Total	256	72	72	400

- (a) Formulate the null hypothesis of no difference in working hours.

<sup>3</sup>These proportions are close to those obtained in a recent Gallup survey.

- (b) Conduct a  $\chi^2$  test of your null hypothesis. Use  $\alpha = .05$ .
- (c) Comment on the form of any departure from the null hypothesis.

13.13 Nausea from air sickness affects some travelers. In a comparative study of the effectiveness of two brands of motion sickness pills, brand A pills were given to 45 persons randomly selected from a group of 90 air travelers, while the other 45 persons were given brand B pills. The following results were obtained.

	Degree of Nausea				Total
	None	Slight	Moderate	Severe	
Brand A	18	17	6	4	45
Brand B	11	14	14	6	45

Do these observations demonstrate that the two brands of pills are significantly different in quality? Test at  $\alpha = .05$ .

- 13.14 Refer to the hardness of mattresses data in the chapter front piece. Confirm that these data establish a difference in the proportions who did not have lower back pain using:
  - (a) The  $\chi^2$  test with level  $\alpha = .01$ .
  - (b) The Z test with level  $\alpha = .01$ . Calculate the P-value.

13.15 A community paper in the Spanish language was delivered to many sites in the San Francisco Bay area. As a check on the circulation numbers

that are important to advertisers, a survey was conducted at four drop sites. The number of papers delivered to each site and the number remaining after 3 days was recorded

	Papers Delivered	Papers Remaining
Site 1	50	17
Site 2	47	12
Site 3	48	7
Site 4	50	21

	Change in Bone Mineral			Total
	Appreciable Loss	Little Change	Appreciable Increase	
Control	38	15	7	60
Therapy	22	32	16	70
Activity	15	30	25	70
Total	75	77	48	200

Do these data indicate that the change in bone mineral varies for the different groups?

- (a) Formulate a null hypothesis of no difference in the proportions of papers taken from the sites.
- (b) Conduct a  $\chi^2$  test of your null hypothesis. Use  $\alpha = .05$ .
- (c) Let  $p_1, p_2, p_3,$  and  $p_4$  denote the probabilities that a paper will be taken from drop sites 1, 2, 3, and 4, respectively. Construct the four individual 95% confidence intervals and plot these intervals.

13.16 Using the data for site 1 and site 3 in Exercise 13.15, make a  $2 \times 2$  table and test  $H_0: p_1 = p_3$  versus  $H_1: p_1 \neq p_3$  at  $\alpha = .05$  using:

- (a) The  $\chi^2$  test.
- (b) The Z test.

13.17 Referring to the data for site 3 and site 4 in Exercise 13.15, make a  $2 \times 2$  table and test  $H_0: p_3 = p_4$  versus  $H_1: p_3 > p_4$  at  $\alpha = .05$ :

- (a) Is there strong evidence that the probability a paper will be taken from the drop site is higher for site 3 than for site 4? Answer by calculating the P-value.
- (b) Construct a 95% confidence interval for  $p_3 - p_4$ .

13.18 Osteoporosis, or a loss of bone minerals, is a common cause of broken bones in the elderly. A researcher on aging conjectures that bone mineral loss can be reduced by regular physical therapy or certain kinds of physical activity. A study is conducted on 200 elderly subjects of approximately the same age divided into control, physical therapy, and physical activity groups. After a suitable period of time, the nature of change in bone mineral content is observed.

### The Following Exercises May Require a Computer

13.19 *Using the computer.* The analysis of a contingency table can be conveniently done on a computer. For an illustration, we present here a MINITAB analysis of the data in Table 6, Example 5.

Data: C13T6.txt

C1: 37 17  
 C2: 24 33  
 C3: 19 20

Dialog box:

Stat > Tables > Chisquare Test  
 Type C1-C3 in Columns containing the table. Click OK.

The output is as follows:

**Chi-Square Test: C1, C2, C3**

**Expected counts are printed below observed counts**

**Chi-Square contributions are printed below expected counts**

	C1	C2	C3	Total
1	37 28.80 2.335	24 30.40 1.347	19 20.80 0.156	80
2	17 25.20 2.668	33 26.60 1.540	20 18.20 0.178	70
<b>Total</b>	<b>54</b>	<b>57</b>	<b>39</b>	<b>150</b>

**Chi-Sq = 8.224, DF = 2, P-Value = 0.016**

- (a) Compare this output with the calculations presented in Example 5.
  - (b) Do Exercise 13.11 on a computer.
  - (c) Do Exercise 13.18 on a computer.
- 13.20 With reference to the sleep data in Table D.10 of the Data Bank, make two categories of snorers on the basis of the last variable: those who responded three or more times a week so their score is coded 4 or 5 and those who responded less than three times a week. Test the equality of proportions for males and females using the  $\chi^2$  test with  $\alpha = .05$ .

#### 4. CONTINGENCY TABLE WITH NEITHER MARGIN FIXED (TEST OF INDEPENDENCE)

When two traits are observed for each element of a random sample, the data can be simultaneously classified with respect to these traits. We then obtain a two-way contingency table in which both sets of marginal totals are random. An illustration was already provided in Example 3. To cite a few other examples: A random sample of employed persons may be classified according to educational attainment and type of occupation; college students may be classified according to the year in college and attitude toward a dormitory regulation; flowering plants may be classified with respect to type of foliage and size of flower.

A typical inferential aspect of cross-tabulation is the study of whether the two characteristics appear to be manifested independently or certain levels of one characteristic tend to be associated or contingent with some levels of another.

#### Example 9 Developing a $\chi^2$ Test of Independence

Analyze the data in Example 3 concerning credit fraud and release of personal information.

#### SOLUTION

The  $2 \times 3$  contingency table of Example 3 is given in Table 12. Here a single random sample of 395 persons, who were notified that personal credit information was released by the institution, were classified according to type of institution and whether or not personal credit was harmed.

**TABLE 12** Type of Organization and Fraudulent Use of Personal Information

	Government	Financial	Other Commercial	Total
Harm	30	24	21	75
No Harm	182	104	34	320
Total	212	128	55	395

Dividing the cell frequencies by the sample size 395, we obtain the relative frequencies shown in Table 13(a). Its row marginal totals .190 and .810 represent the sample proportions of cases where credit harm was inflicted and where it was not, respectively. Likewise, the column marginal totals show

**TABLE 13(a)** Proportion of Observations in Each Cell

	Government	Financial	Other Commercial	Total
Harm	.076	.061	.053	.190
No Harm	.461	.263	.086	.810
Total	.537	.324	.139	1.000

the sample proportions for the three types of institutions. For instance, Government was the source of proportion .537, or over half of the releases.

Imagine a classification of the entire population. The unknown population proportions (i.e., the probabilities of the cells) are represented by the entries in Table 13(b), where the suffixes H and N stand for Harm and No Harm, and 1, 2, and 3 refer to “Government,” “Financial,” and “Other Commercial” institutions. Table 13(b) is the population analogue of Table 13(a), which shows the sample proportions. For instance

$$\text{Cell probability } p_{H1} = P(\text{Government and Harm})$$

$$\text{Row marginal probability } p_H = P(\text{Harm})$$

$$\text{Column marginal probability } p_1 = P(\text{Government})$$

**TABLE 13(b)** Probability of Each Cell

	Government	Financial	Other Commercial	Row Marginal Probability
Harm	$p_{H1}$	$p_{H2}$	$p_{H3}$	$p_H$
No Harm	$p_{N1}$	$p_{N2}$	$p_{N3}$	$p_N$
Column marginal probability	$p_1$	$p_2$	$p_3$	1

We are concerned with testing the null hypothesis that the two classifications are independent. Recall from Chapter 4 that the probability of the intersection of independent events is the product of their probabilities. Thus, the independence of the two classifications means that  $p_{H1} = p_H p_1$ ,  $p_{H2} = p_H p_2$ , and so on.

Therefore, the null hypothesis of independence can be formalized as

$$H_0 : \text{Each cell probability is the product of the corresponding pair of marginal probabilities.}$$

To construct a  $\chi^2$  test, we need to determine the expected frequencies. Under  $H_0$ , the expected cell frequencies are

$$\begin{array}{ccc} 395 p_H p_1 & 395 p_H p_2 & 395 p_H p_3 \\ 395 p_N p_1 & 395 p_N p_2 & 395 p_N p_3 \end{array}$$

These involve the unknown marginal probabilities that must be estimated from the data. Referring to Table 12, we calculate the estimates as

$$\hat{p}_H = \frac{75}{395} \quad \hat{p}_N = \frac{320}{395}$$

$$\hat{p}_1 = \frac{212}{395} \quad \hat{p}_2 = \frac{128}{395} \quad \hat{p}_3 = \frac{56}{395}$$

Then, the expected frequency in the first cell is estimated as

$$395\hat{p}_H\hat{p}_1 = 395 \times \frac{75}{395} \times \frac{102}{395}$$

$$= \frac{75 \times 212}{395} = 40.25$$

Notice that the expected frequency for each cell of Table 12 is of the form

$$\frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

Table 14(a) presents the observed cell frequencies along with the expected frequencies shown in parentheses. The quantities  $(O - E)^2/E$  and the  $\chi^2$  statistic are then calculated in Table 13(b).

**TABLE 14(a)** The Observed and Expected Cell Frequencies for the Credit Fraud Data in Table 12

	Government	Financial	Other Commercial
Harm	30 (40.25)	24 (24.30)	21 (10.44)
No Harm	182 (171.75)	104 (103.70)	34 (44.56)

**TABLE 14(b)** The Values of  $(O - E)^2/E$

	Government	Financial	Other Commercial	Total
Harm	2.612	.004	10.672	
No Harm	.612	.001	2.501	
				16.402 = $\chi^2$

Having calculated the  $\chi^2$  statistic, it now remains to determine its d.f. by invoking the properties we stated in Section 2. Because we have a single

random sample, only the property (b) is relevant to this problem. Since  $p_H + p_N = 1$  and  $p_1 + p_2 + p_3 = 1$ , we have really estimated  $1 + 2 = 3$  parameters. Hence,

$$\begin{aligned} \text{d.f. of } \chi^2 &= (\text{No. of cells}) - 1 - (\text{No. of parameters estimated}) \\ &= 6 - 1 - 3 \\ &= 2 \end{aligned}$$

We choose level of significance  $\alpha = .05$  and the tabulated upper 5% point of  $\chi^2$  with d.f. = 2 is 5.99. Because the observed  $\chi^2$  is larger than the tabulated value, the null hypothesis of independence is rejected at  $\alpha = .05$ . In fact, it would be rejected even for  $\alpha = .01$ .

Inspection of Table 14(b) reveals that moderately large contributions from (i) the cell where Government released personal data has a smaller than expected number of cases where credit is harmed and (ii) the cell where Other Commercial institutions have smaller than expected number of cases where credit is not harmed. However, a gigantic contribution to  $\chi^2$  comes from the cell where personal data released by Other Commercial institutions has much larger than expected number of cases where credit was harmed.

In all cases where you are notified that your personal information has been compromised, you need to check your credit reports and carefully read credit card statements. This is particularly true if other commercial institutions are involved.

From our analysis of the contingency table in Example 9, the  $\chi^2$  test of independence in a general  $r \times c$  contingency table is readily apparent. In fact, it is much the same as the test of homogeneity described in Section 3. The expected cell frequencies are determined in the same way—namely,

$$\text{Expected cell frequency} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

and the test statistic is again

$$\chi^2 = \sum_{\text{cells}} \frac{(O - E)^2}{E}$$

With regard to the d.f. of  $\chi^2$  in the present case, we initially have  $rc - 1$  d.f. because there are  $rc$  cells into which a single random sample is classified. From this, we must subtract the number of estimated parameters. This number is  $(r - 1) + (c - 1)$  because there are  $r - 1$  parameters among the row marginal probabilities and  $c - 1$  parameters among the column marginal probabilities. Therefore,

$$\begin{aligned} \text{d.f. of } \chi^2 &= rc - 1 - (r - 1) - (c - 1) \\ &= rc - r - c + 1 \\ &= (r - 1)(c - 1) \\ &= (\text{No. of rows} - 1) \times (\text{No. of columns} - 1) \end{aligned}$$

which is identical to the d.f. of  $\chi^2$  for testing homogeneity. In summary, the  $\chi^2$  test statistic, its d.f., and the rejection region for testing independence are the same as when testing homogeneity. It is only the statement of the null hypothesis that is different between the two situations.

### The Null Hypothesis of Independence

$H_0$  : Each cell probability equals the product of the corresponding row and column marginal probabilities.

### SPURIOUS DEPENDENCE

When the  $\chi^2$  test leads to a rejection of the null hypothesis of independence, we conclude that the data provide evidence of a **statistical association** between the two characteristics. However, we must refrain from making the hasty interpretation that these characteristics are directly related. A claim of casual relationship must draw from common sense, which statistical evidence must not be allowed to supersede.

Two characteristics may appear to be strongly related due to the common influence of a third factor that is not included in the study. In such cases, the dependence is called a **spurious dependence**. For instance, if a sample of individuals is classified in a  $2 \times 2$  contingency table according to whether or not they are heavy drinkers and whether or not they suffer from respiratory trouble, we would probably find a high value for  $\chi^2$  and conclude that a strong statistical association exists between drinking habit and lung condition. But the reason for the association may be that most heavy drinkers are also heavy smokers and the smoking habit is a direct cause of respiratory trouble. This discussion should remind the reader of a similar warning given in Chapter 3 regarding the interpretation of a correlation coefficient between two sets of measurements. In the context of contingency tables, examples of spurious dependence are sometimes called **Simpson's paradox**, which is discussed in Chapter 3, Section 2.

### Exercises

- 13.21 Applicants for public assistance are allowed an appeals process when they feel unfairly treated. At such a hearing, the applicant may choose self-representation or representation by an attorney. The appeal may result in an increase, decrease, or no change of the aid recommendation. Court records of 320 appeals cases provided the following data.

Type of Representation	Amount of Aid		
	Increased	Unchanged	Decreased
Self	59	108	17
Attorney	70	63	3

- Test the null hypothesis that the appeals decision and the type of representation are independent. Test at  $\alpha = .05$ .
- 13.22 A consultant to all kinds of retailers suggests that they have plenty of baskets available for customers. He bases his suggestion on data collected by watching video tapes from hidden cameras. Suppose that out of 200 customers, 80 picked up a basket when they entered the store. Among those who picked up baskets 60 persons made purchases while only 41 of those without baskets made a purchase.

- (a) Conduct a  $\chi^2$  test of independence between purchasing and the decision to pick up a basket. Use  $\alpha = .05$ .
- (b) The consultant may be confusing association with cause and effect. For instance, do you think the decision to pick up a basket has anything to do with the intent to purchase when a person enters the store? Comment.

13.23 A survey was conducted by sampling 400 persons who were questioned regarding union membership and attitude toward decreased national spending on social welfare programs. The cross-tabulated frequency counts are presented.

	Support	Indifferent	Opposed	Total
Union	112	36	28	176
Nonunion	84	68	72	224
Total	196	104	100	400

Can these observed differences be explained by chance or are attitude and membership status associated?

13.24 A survey was conducted to study people's attitude toward television programs that show violence. A random sample of 1200 adults was selected and classified according to gender and response to the question: Do you think there is a link between violence on TV and crime?

	Response		
	Yes	No	Not Sure
Male	361	228	17
Female	433	141	20

Do the survey data show a significant association between attitude and gender?

13.25 In a study of factors that regulate behavior, three kinds of subjects are identified: overcontrollers,

average controllers, and undercontrollers, with the first group being most inhibited. Each subject is given the routine task of filling a box with buttons and all subjects are told they can stop whenever they wish. Whenever a subject indicates he or she wishes to stop, the experimenter asks, "Don't you really want to continue?" The number of subjects in each group who stop and the number who continue are given in the following table.

Controller	Continue	Stop	Total
Over	9	9	18
Average	8	12	20
Under	3	14	17
Total	20	35	55

Are controller group and continue/stop decision associated?

13.26 Out of 120 members of a club who responded to a survey, 80 said that the golfing facilities were influential, 53 said the dining facilities were influential, and 25 said both were influential in their decision to join the club.

- (a) Construct a two-way table of frequencies with two categories influential/not influential for each facility, golf and dining.
- (b) Formulate the null hypothesis.
- (c) Conduct a  $\chi^2$  test of your null hypothesis. Use  $\alpha = .05$ .
- (d) Comment on the form of any departure from the null hypothesis.

**The Following Exercises May Require a Computer**

13.27 *Using the computer.* The analysis of a contingency table becomes tedious especially when the size of the table is large. Using a computer makes the task quite easy. We illustrate this by using MINITAB to analyze the data in Table 12, Example 9.



Data: C13T12.txt

C1: 30 182

C2: 24 104

C3: 21 34

Dialog box:

Stat > Tables > Chisquare Test

Type C1-C3 in Columns containing the table. Click OK.

Chi-Square contributions are printed below expected counts

	C1	C2	C3	Total
1	30	24	21	75
	40.25	24.30	10.44	
	2.612	0.004	10.672	
2	182	104	34	320
	171.75	103.70	44.56	
	0.612	0.001	2.501	
Total	212	128	55	395

Chi-Sq = 16.402, DF = 2,

P-Value = 0.000

The output is as follows:

Chi-Square Test: C1, C2, C3

Expected counts are printed below observed counts

(a) Compare this output with the calculations presented in Example 9.

(b) Do Exercise 13.24 on a computer.

13.28 Do Exercise 13.26 on a computer.

## USING STATISTICS WISELY

- Moderately large sample sizes are required to detect differences in proportions. Usually, 50 to 100 observations from each population are needed.
- Although the  $\chi^2$  test statistic is the same for testing independence and for testing equality of proportions, you should be clear which null hypothesis you are testing. When you sample from separate populations, the  $\chi^2$  test concerns the equality of proportions. If a single sample is cross-classified according to two characteristics, the  $\chi^2$  test concerns independence.
- Don't routinely apply the inference procedures for comparing proportions when it is obvious that the outcomes have a time order dependence.

## KEY IDEAS AND FORMULAS

The term **categorical data** refers to observations that are only classified into categories so the data consist of frequency counts for the categories. When the frequency counts arise because observations are classified according to two or more characteristics, they are called **cross-tabulated data** or a **contingency table**. A **chi-square ( $\chi^2$ ) statistic** compares the observed frequencies with those expected under a null hypothesis.

When a chi-square test of independence leads to the rejection of the null hypothesis, we say we have established a **statistical association**. There is no assertion

of a causal relation as it may be a **spurious dependence** caused by a third variable. **Simpson's paradox** is an example of spurious dependence caused when two categorical data sets collected from very different populations are combined.

### Pearson's $\chi^2$ Test for Goodness of Fit

Data: Observed cell frequencies  $n_1, \dots, n_k$  from a random sample of size  $n$  classified into  $k$  cells.

The null hypothesis specifies the cell probabilities

$$H_0: p_1 = p_{10}, \dots, p_k = p_{k0}$$

Test statistic

$$\chi^2 = \sum_{\text{cells}} \frac{(n_i - np_{i0})^2}{np_{i0}} \quad \text{d.f.} = k - 1$$

$$\text{Rejection region } \chi^2 \geq \chi_{\alpha}^2$$

### $\chi^2$ Test of Homogeneity in an $r \times c$ Contingency Table

Data: Independent random samples from  $r$  populations, each sample classified in  $c$  response categories.

Null hypothesis: In each response category, the probabilities are equal for all the populations.

Test statistic

$$\chi^2 = \sum_{\text{cells}} \frac{(O - E)^2}{E} \quad \text{d.f.} = (r - 1)(c - 1)$$

where for each cell

$O$  = Observed cell frequency

$$E = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

$$\text{Rejection region } \chi^2 \geq \chi_{\alpha}^2$$

### $\chi^2$ Test of Independence in an $r \times c$ Contingency Table

Data: A random sample of size  $n$  is simultaneously classified with respect to two characteristics, one has  $r$  categories and the other  $c$  categories.

Null hypothesis: The two classifications are independent; that is, each cell probability is the product of the row and column marginal probabilities.

Test statistic and rejection region: Same as when testing homogeneity.

### Limitation

All inference procedures of this chapter require large samples. The  $\chi^2$  tests are appropriate if no expected cell frequency is too small ( $\geq 5$  is normally required).

## TECHNOLOGY

### *Conducting a $\chi^2$ test*

#### MINITAB

### *Conducting a $\chi^2$ test*

We illustrate the calculation of the  $\chi^2$  statistic with an example. Enter the counts

30	42	28
9	10	31

in columns 1 to 3. Select the MINITAB commands

**Stat > Tables > Chisquare Test.**  
 Type *C1–C3* in **columns containing the table.**  
 Click **OK.**

#### EXCEL

### *Calculating a $\chi^2$ statistic*

Enter observed values in a rectangular range of cells.

Enter the expected values in another rectangular range of cells.

Select **Insert**, then **Function**. Select **Statistical**, and then **CHITEST**.  
 Click **OK**.

With the cursor in the textbox for **Actual\_range**, highlight the observed values.

With the cursor in the textbox for **Expected\_range**, highlight the expected values.

Click **OK**.

The software will return the  $P$ -value.

#### TI-84/-83 PLUS

### *Calculating a $\chi^2$ statistic*

Enter the observed counts and expected values as matrices.

Select **Matrix** and then **EDIT**.

Press **Enter** and enter the number of rows, the number of columns, and the entries for **[A]**, the observed values.

Press **2nd Quit** and select **Matrix**, then **EDIT** again.

Select **Matrix [B]**, press **Enter** and enter the number of rows, the number of columns, and the expected values. Press **2nd Quit** again.

Select **STAT** then **TESTS** and then **>  $\chi^2$  test**.

Arrow down to select **Calculate**. Enter the appropriate matrix names (matrices [A] and [B] are the defaults) and then press **Enter**.

The software returns the value of the  $\chi^2$  statistic, the  $P$ -value, and the degrees of freedom.

## 5. REVIEW EXERCISES

- 13.29 To examine the quality of a random number generator, frequency counts of the individual integers are recorded from an output of 500 integers. The concept of randomness implies that the integers 0, 1, . . . , 9 are equally likely. Based on the observed frequency counts, would you suspect any bias of the random number generator? Answer by performing the  $\chi^2$  test.

Integer	0	1	2	3	4	5	6	7	8	9	Total
Frequency	41	58	51	61	39	56	45	35	62	52	500

- 13.30 The following record shows a classification of 41,208 births in Wisconsin (courtesy of Professor Jerome Klotz). Test the goodness of fit of the model that births are uniformly distributed over all 12 months of the year. Use  $\alpha = .01$ .

Jan.	3,478	July	3,476
Feb.	3,333	Aug.	3,495
Mar.	3,771	Sept.	3,490
Apr.	3,542	Oct.	3,331
May	3,479	Nov.	3,188
June	3,304	Dec.	<u>3,321</u>
		Total	41,208

- 13.31 The following table records the observed number of births at a hospital in four consecutive quarterly periods.

Quarters	Jan.–March	April–June	July–Sept.	Oct.–Dec.
Number of births	55	29	26	41

It is conjectured that twice as many babies are born during the Jan.–March quarter than any of the other three quarters. At  $\alpha = .10$ , test if these data strongly contradict the stated conjecture.

- 13.32 A large midwestern university allows multiple majors and most educators appreciate that double majors are warranted in some circumstances. Recently, however, there is more concern about what a major means if a person has too many. The data on majors for graduating seniors in the College of Letters and Sciences, and the results for an earlier academic year are

	Number of majors		
	1	2	3
01–02 Year	2327	834	57
07–08 Year	2356	1059	88

Treating these data as a random sample of graduating seniors at similar institutions,

- (a) perform a  $\chi^2$  test to detect a difference in the populations for number of majors. Take  $\alpha = .05$ .  
 (b) Let  $p_1$  be the probability a 2007–2008 graduate has multiple majors and let  $p_2$  be the probability for 2001–2002. Perform a  $\chi^2$  test for a difference. Take  $\alpha = .05$ .

- 13.33 Refer to Exercise 3.48 and the data concerning a vaccine for type B hepatitis.

	Hepatitis	No Hepatitis	Total
Vaccinated	11	538	549
Not vaccinated	70	464	534
Total	81	1002	1083

Do these data indicate that there is a different rate of incidence of hepatitis between the vaccinated and nonvaccinated participants? Use the  $\chi^2$  test for homogeneity in a contingency table.

- 13.34 Refer to the data in Exercise 13.33.
- Use the  $Z$  test for testing the equality of two population proportions with a two-sided alternative. Verify the relation  $\chi^2 = Z^2$  by comparing their numerical values.
  - If the alternative is that the incidence rate is lower for the vaccinated group, which of the two tests should be used?

- 13.35 To compare the effectiveness of four drugs in relieving postoperative pain, an experiment was done by randomly assigning 195 surgical patients to the drugs under study. Recorded here are the number of patients assigned to each drug and the number of patients who were free of pain for a period of five hours.

	Free of Pain	No. of Patients Assigned
Drug 1	23	53
Drug 2	30	47
Drug 3	19	51
Drug 4	29	44

- Make a  $4 \times 2$  contingency table showing the counts of patients who were free of pain and those who had pain, and test the null hypothesis that all four drugs are equally effective. (Use  $\alpha = .05$ .)
  - Let  $p_1, p_2, p_3,$  and  $p_4$  denote the population proportions of patients who would be free of pain under the use of drugs 1, 2, 3, and 4, respectively. Calculate a 90% confidence interval for each of these probabilities individually.
- 13.36 Using the data for drugs 1 and 3 in Exercise 13.35, make a  $2 \times 2$  contingency table and test  $H_0: p_1 = p_3$  versus  $H_1: p_1 \neq p_3$  at  $\alpha = .05$  employing:
- The  $\chi^2$  test.
  - The  $Z$  test.
- 13.37 Refer to the data for drugs 3 and 4 in Exercise 13.35.

- Is there strong evidence that drug 4 is more effective in controlling postoperative pain than drug 3? Answer by calculating the  $P$ -value.
- Construct a 95% confidence interval for the difference  $p_4 - p_3$ .

- 13.38 In a study on the effect of diet and lifestyle on heart disease, 96 patients with severe coronary blockage were randomly assigned, 49 to an experimental group and 47 to a control group. The patients in the experimental group had a low-fat vegetarian diet, regular exercise, and stress-management training, whereas those in the control group had a low-fat diet and moderate exercise. The condition of their coronary blockage was monitored throughout the study period, and the following results were noted.

Group	Coronary Blockage			Total
	Worsened	No Change	Improved	
Experimental	4	8	37	49
Control	8	25	14	47
				96

Analyze the data to determine if the changes in coronary blockage were significantly different between the two groups of patients.

- 13.39 Based on interviews of couples seeking divorces, a social worker compiles the following data related to the period of acquaintanceship before marriage and the duration of marriage.

Acquaintanceship before Marriage	Duration of Marriage		Total
	$\leq 4$ years	$> 4$ years	
Under $\frac{1}{2}$ year	11	8	19
$\frac{1}{2} - 1\frac{1}{2}$ years	28	24	52
Over $1\frac{1}{2}$ years	21	19	40
Total	60	51	111

Perform a test to determine if the data substantiate an association between the stability of a marriage and the period of acquaintanceship prior to marriage.

- 13.40 By polling a random sample of 350 undergraduate students, a campus press obtains the following frequency counts regarding student attitude toward a proposed change in dormitory regulations.

	Favor	Indifferent	Oppose	Total
Male	95	72	19	186
Female	53	79	32	164
Total	148	151	51	350

Are attitude toward the proposal and gender associated?

- 13.41 In a genetic study of chromosome structures, 132 individuals are classified according to the type of structural chromosome aberration and carriers in their parents. The following counts are obtained.

Type of Aberration	Carrier		Total
	One Parent	Neither Parent	
Presumably innocuous	27	20	47
Substantially unbalanced	36	49	85
Total	63	69	132

Test the null hypothesis that type of aberration is independent of parental carrier.

- 13.42 A random sample of 130 business executives was classified according to age and the degree of risk aversion as measured by a psychological test.

Age	Degree of Risk Aversion			Total
	Low	Medium	High	
Below 45	14	22	7	43
45–55	16	33	12	61
Over 55	4	15	7	26
				130

Do these data demonstrate an association between risk aversion and age?

- 13.43 *Pooling contingency tables can produce spurious association.* A large organization is being investigated to determine if its recruitment is sex-biased. Tables 14 and 15, respectively, show the classification of applicants for secretarial and for sales positions according to gender and result of interview. Table 16 is an aggregation of the corresponding entries of Table 14 and Table 15.

**TABLE 14** Secretarial Positions

	Offered	Denied	Total
Male	25	50	75
Female	75	150	225
Total	100	200	300

**TABLE 15** Sales Positions

	Offered	Denied	Total
Male	150	50	200
Female	75	25	100
Total	225	75	300

**TABLE 16** Secretarial and Sales Positions

	Offered	Denied	Total
Male	175	100	275
Female	150	175	325
Total	325	275	600

- Verify that the  $\chi^2$  statistic for testing independence is zero for each of the data sets given in Tables 14 and 15.
- For the pooled data given in Table 16, compute the value of the  $\chi^2$  statistic and test the null hypothesis of independence.
- Explain the paradoxical result that there is no sex bias in any job category, but the combined data indicate sex discrimination.

# Analysis of Variance (ANOVA)

1. Introduction
2. Comparison of Several Treatments—  
The Completely Randomized Design
3. Population Model and Inferences  
for a Completely Randomized Design
4. Simultaneous Confidence Intervals
5. Graphical Diagnostics and Displays  
to Supplement ANOVA
6. Randomized Block Experiments  
for Comparing  $k$  Treatments
7. Review Exercises

---

## *Which Brand HDTV Has the Clearest Picture?*



A good experimental design is to collect samples of several HDTVs of each brand and measure their picture clarity. The statistical technique called “analysis of variance” enables us to verify differences among the brands.  
© Ed Lallo/Index Stock Imagery.

---



## 1. INTRODUCTION

In Chapter 10, we introduced methods for comparing two population means. When several means must be compared, more general methods are required. We now become acquainted with the powerful technique called **analysis of variance (ANOVA)** that allows us to analyze and interpret observations from several populations. This versatile statistical tool partitions the total variation in a data set according to the sources of variation that are present. In the context of comparing  $k$  population means, the two sources of variation are (1) differences **between** means or **treatments** and (2) **within** population variation (**error**). We restrict our discussion to this case, although ANOVA techniques apply to much more complex situations.

In this chapter, you will learn how to test for differences among several means and to make confidence statements about pairs of means.

## 2. COMPARISON OF SEVERAL TREATMENTS — THE COMPLETELY RANDOMIZED DESIGN

It is usually more expedient in terms of both time and expense to simultaneously compare several treatments than it is to conduct several comparative trials two at a time. The term **completely randomized design** is synonymous with independent random sampling from several populations when each population is identified as the population of responses under a particular treatment. Let treatment 1 be applied to  $n_1$  experimental units, treatment 2 to  $n_2$  units, . . . , treatment  $k$  to  $n_k$  units. In a completely randomized design,  $n_1$  experimental units selected at random from the available collection of  $n = n_1 + n_2 + \cdots + n_k$  units are to receive treatment 1,  $n_2$  units randomly selected from the remaining units are to receive treatment 2, and proceeding in this manner, treatment  $k$  is to be applied to the remaining  $n_k$  units. The special case of this design for a comparison of  $k = 2$  treatments has already been discussed in Section 2 of Chapter 10. The data structure for the response measurements can be represented by the format shown in Table 1, where  $y_{ij}$  is the  $j$ th observation on treatment  $i$ . The summary statistics appear in the last two columns.

Before proceeding with the general case of  $k$  treatments, it would be instructive to explain the reasoning behind the analysis of variance and the associated calculations in terms of a numerical example.

### Example 1 The Structure of Data from an Experiment for Comparing Four Means

Four models of high-end earbuds will be compared for sound reproduction. There are 5 sets of earbud A, 4 of earbud B, 7 of earbud C, and 6 of earbud D available for testing. The quality of sound reproduction can be determined objectively by measuring audio signals received by a robot head wearing earbuds and then comparing them with the known signal wave that was sent. Quantitatively, the measure of sound distortion called total harmonic distortion is an overall measure of the discrepancy, in percent. Because the values

**TABLE 1** Data Structure for the Completely Randomized Design with  $k$  Treatments

	Observations	Mean	Sum of Squares
Treatment 1	$y_{11}, y_{12}, \dots, y_{1n_1}$	$\bar{y}_1$	$\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2$
Treatment 2	$y_{21}, y_{22}, \dots, y_{2n_2}$	$\bar{y}_2$	$\sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2$
	$\vdots$	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$
Treatment $k$	$y_{k1}, y_{k2}, \dots, y_{kn_k}$	$\bar{y}_k$	$\sum_{j=1}^{n_k} (y_{kj} - \bar{y}_k)^2$
Grand mean $\bar{y}$	$= \frac{\text{Sum of all observations}}{n_1 + n_2 + \dots + n_k}$	$= \frac{n_1\bar{y}_1 + \dots + n_k\bar{y}_k}{n_1 + \dots + n_k}$	

are substantially below 1% for high quality earbuds, we give the values for distortion in hundredths of a percent so 10 is .1% and so on. Suppose the test results for sound distortion produce the data in Table 2.

Two questions immediately come to mind. Does any significant difference exist among the mean distortions obtained using the four brands of earbuds? Can we establish confidence intervals for the mean differences between brands?

**TABLE 2** Sound Distortion Obtained with Four Brands of Earbuds

Coating	Observations	Mean	Sum of Squares
A	10, 15, 8, 12, 15	$\bar{y}_1 = 12$	$\sum_{j=1}^5 (y_{1j} - \bar{y}_1)^2 = 38$
B	14, 18, 21, 15	$\bar{y}_2 = 17$	$\sum_{j=1}^4 (y_{2j} - \bar{y}_2)^2 = 30$
C	17, 16, 14, 15, 17, 15, 18	$\bar{y}_3 = 16$	$\sum_{j=1}^7 (y_{3j} - \bar{y}_3)^2 = 12$
D	12, 15, 17, 15, 16, 15	$\bar{y}_4 = 15$	$\sum_{j=1}^6 (y_{4j} - \bar{y}_4)^2 = 14$
	Grand mean $\bar{y}$	$= 15$	

An analysis of the results essentially consists of decomposing the observations into contributions from different sources. We reason that the deviation of an individual observation from the grand mean,  $y_{ij} - \bar{y}$ , is partly due to differences among the mean qualities of the brands and partly due to random



Thus, the sum of squares due to differences in the treatment means, also called the **treatment sum of squares**, is given by

$$\text{Treatment sum of squares} = \sum_{i=1}^4 n_i (\bar{y}_i - \bar{y})^2 = 68$$

The last array consists of the entries  $y_{ij} - \bar{y}_i$  that are the deviations of individual observations from the corresponding treatment mean. These deviations reflect inherent variabilities in the material, fabrication, and the measuring device and are called the **residuals**. The overall variation due to random errors is measured by the sum of squares of all these residuals

$$(-2)^2 + 3^2 + (-4)^2 + \cdots + 1^2 + 0^2 = 94$$

Thus, we obtain

$$\text{Error sum of squares} = \sum_{i=1}^4 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = 94$$

The double summation indicates that the elements are summed within each row and then over different rows. Alternatively, referring to the last column in Table 2, we obtain

$$\begin{aligned} \text{Error sum of squares} &= \sum_{j=1}^5 (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^4 (y_{2j} - \bar{y}_2)^2 \\ &\quad + \sum_{j=1}^7 (y_{3j} - \bar{y}_3)^2 + \sum_{j=1}^6 (y_{4j} - \bar{y}_4)^2 \\ &= 38 + 30 + 12 + 14 = 94 \end{aligned}$$

Finally, the deviations of individual observations from the grand mean  $y_{ij} - \bar{y}$  are given by the array

$$\text{Deviations} = \begin{bmatrix} -5 & 0 & -7 & -3 & 0 & & & \\ -1 & 3 & 6 & 0 & & & & \\ 2 & 1 & -1 & 0 & 2 & 0 & 3 & \\ -3 & 0 & 2 & 0 & 1 & 0 & & \end{bmatrix}$$

The total variation present in the data is measured by the sum of squares of all these deviations.

$$\begin{aligned} \text{Total sum of squares} &= \sum_{i=1}^4 \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \\ &= (-5)^2 + 0^2 + (-7)^2 + \cdots + 0^2 \\ &= 162 \end{aligned}$$

Note that the total sum of squares is the sum of the treatment sum of squares and the error sum of squares.

It is time to turn our attention to another property of this decomposition, the degrees of freedom associated with the sums of squares. In general terms:

$$\left( \begin{array}{l} \text{Degrees of} \\ \text{freedom} \\ \text{associated with a} \\ \text{sum of squares} \end{array} \right) = \left( \begin{array}{l} \text{Number of} \\ \text{elements} \\ \text{whose squares} \\ \text{are summed} \end{array} \right) - \left( \begin{array}{l} \text{Number of linear} \\ \text{constraints} \\ \text{satisfied by the} \\ \text{elements} \end{array} \right)$$

In our present example, the treatment sum of squares is the sum of four terms  $n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2 + n_3(\bar{y}_3 - \bar{y})^2 + n_4(\bar{y}_4 - \bar{y})^2$ , where the elements satisfy the single constraint

$$n_1(\bar{y}_1 - \bar{y}) + n_2(\bar{y}_2 - \bar{y}) + n_3(\bar{y}_3 - \bar{y}) + n_4(\bar{y}_4 - \bar{y}) = 0$$

This equality holds because the grand mean  $\bar{y}$  is a weighted average of the treatment means, or

$$\bar{y} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2 + n_3\bar{y}_3 + n_4\bar{y}_4}{n_1 + n_2 + n_3 + n_4}$$

Consequently, the number of degrees of freedom associated with the treatment sum of squares is  $4 - 1 = 3$ . To determine the degrees of freedom for the error sum of squares, we note that the entries  $y_{ij} - \bar{y}_i$  in each row of the residual array sum to zero and there are 4 rows. The number of degrees of freedom for the error sum of squares is then  $(n_1 + n_2 + n_3 + n_4) - 4 = 22 - 4 = 18$ . Finally, the number of degrees of freedom for the total sum of squares is  $(n_1 + n_2 + n_3 + n_4) - 1 = 22 - 1 = 21$ , because the 22 entries  $(y_{ij} - \bar{y})$  whose squares are summed satisfy the single constraint that their total is zero. Note that the degrees of freedom for the total sum of squares is the sum of the degrees of freedom for treatment and error.

We summarize the calculations thus far in Table 3.

**TABLE 3** ANOVA Table for Distortion Data

Source	Sum of Squares	d.f.
Treatment	68	3
Error	94	18
Total	162	21

Guided by this numerical example, we now present the general formulas for the analysis of variance for a comparison of  $k$  treatments using the data structure given in Table 1. Beginning with the basic decomposition

$$(y_{ij} - \bar{y}) = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$$

and squaring each side of the equation, we obtain

$$(y_{ij} - \bar{y})^2 = (\bar{y}_i - \bar{y})^2 + (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \bar{y})(y_{ij} - \bar{y}_i)$$

When summed over  $j = 1, \dots, n_i$  the last term on the right-hand side of this equation reduces to zero due to the relation  $\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i) = 0$ . Therefore, summing each side of the preceding relation over  $j = 1, \dots, n_i$  and  $i = 1, \dots, k$  provides the decomposition

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

$\uparrow$   
**Total SS**  
 d.f. =  $\sum_{i=1}^k n_i - 1$

$\uparrow$   
**Treatment SS**  
 d.f. =  $k - 1$

$\uparrow$   
**Residual SS  
or error SS**  
 d.f. =  $\sum_{i=1}^k n_i - k$

It is customary to present the decomposition of the sum of squares and the degrees of freedom in a tabular form called the **analysis of variance table**, abbreviated as **ANOVA table**. This table contains the additional column for the **mean square** associated with a component, which is defined as

$$\text{Mean square} = \frac{\text{Sum of squares}}{\text{d.f.}}$$

The ANOVA table for comparing  $k$  treatments appears in Table 4.

**TABLE 4** ANOVA Table for Comparing  $k$  Treatments

Source	Sum of Squares	d.f.	Mean Square
Treatment	$SS_T = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$	$k - 1$	$MS_T = \frac{SS_T}{k - 1}$
Error	$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$\sum_{i=1}^k n_i - k$	$MSE = \frac{SSE}{\sum_{i=1}^k n_i - k}$
Total	$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$\sum_{i=1}^k n_i - 1$	

### GUIDE TO HAND CALCULATION

When performing an ANOVA on a calculator, it is convenient to express the sums of squares in an alternative form. These employ the treatment totals

$$T_i = \sum_{j=1}^{n_i} y_{ij} = \text{Sum of all responses under treatment } i$$

$$T = \sum_{i=1}^k T_i = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} = \text{Sum of all observations}$$

to calculate the sums of squares:

$$\text{Total SS} = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{n} \quad \text{where} \quad n = \sum_{i=1}^k n_i$$

$$\text{SS}_T = \sum_{i=1}^k \frac{T_i^2}{n_i} - \frac{T^2}{n}$$

$$\text{SSE} = \text{Total SS} - \text{SS}_T$$

Notice that the SSE can be obtained by subtraction.

#### Example 2 Calculating Sums of Squares Using the Alternative Formulas

Obtain the Total SS,  $\text{SS}_T$ , and SSE for the data in Example 1 using the alternative form of calculation.

**SOLUTION**

$$\begin{aligned} T_1 &= 10 + 15 + 8 + 12 + 15 = 60 & n_1 &= 5 \\ T_2 &= 14 + 18 + 21 + 15 = 68 & n_2 &= 4 \\ T_3 &= 17 + 16 + 14 + 15 + 17 + 15 + 18 = 112 & n_3 &= 7 \\ T_4 &= 12 + 15 + 17 + 15 + 16 + 15 = 90 & n_4 &= 6 \end{aligned}$$

and

$$\begin{aligned} T &= T_1 + T_2 + T_3 + T_4 & n &= n_1 + n_2 + n_3 + n_4 \\ &= 60 + 68 + 112 + 90 = 330 & &= 5 + 4 + 7 + 6 = 22 \end{aligned}$$

Since

$$\begin{aligned} \sum_{i=1}^4 \sum_{j=1}^{n_i} y_{ij}^2 &= (10)^2 + (15)^2 + \cdots + (16)^2 + (15)^2 = 5112 \\ \text{Total SS} &= 5112 - \frac{(330)^2}{22} = 162 \\ \text{SS}_T &= \frac{(60)^2}{5} + \frac{(68)^2}{4} + \frac{(112)^2}{7} + \frac{(90)^2}{6} - \frac{(330)^2}{22} = 68 \\ \text{SSE} &= \text{Total SS} - \text{SS}_T = 162 - 68 = 94 \end{aligned}$$

**Exercises**

14.1 Water collected in a single bottle from a river is divided into eight specimens. Two specimens are randomly selected and sent to Lab A, two to Lab B, two to Lab C and two to Lab D. The amount of heavy metals (ppm) is measured for each specimen. Calling each lab a treatment, suppose the data are

Treatment	Observations
A	5 9
B	8 4
C	4 2
D	7 9

- (a) Obtain the arrays that show a decomposition for the observations.
- (b) Find the sum of squares for each array.
- (c) Determine the degrees of freedom for each sum of squares.
- (d) Summarize by an ANOVA table.

14.2 Subjects must press a button when they hear a signal. The three treatments are three different intensities of the signal. The time elapsed between presentation of the signal and when the button is pushed, are recorded in hundredths of a second. Suppose the data are

Treatment	Observations
1	35, 24, 28, 21
2	19, 14, 14, 13
3	21, 16, 21, 14

- (a) Obtain the arrays that show a decomposition for the observations.
- (b) Find the sum of squares for each array.
- (c) Determine the degrees of freedom for each sum of squares.
- (d) Summarize by an ANOVA table.

14.3 It was decided to vary the experiment in Exercise 14.1 by dropping Lab D. Water from a single collection bottle was divided into eleven specimens. The amount of heavy metals (ppm) is measured for each specimen. Suppose the data are as follows and repeat Exercise 14.1.

Treatment	Observations
A	7 5 4 4
B	6 1 2
C	2 1 0 1

14.4 The abilities of six different brand athletic socks to wick moisture are rated. The rating is based on the time to reach 1% moisture content after being subjected to a gentle mist. Use the relations for sums of squares and d.f. to complete the following ANOVA table:

Source	Sum of Squares	d.f.
Treatment	34	5
Error		
Total	92	25

14.5 Spots cannot always be removed by dry cleaning. Suppose the records from four different dry cleaning establishments yield the following data on number of unremovable spots per day.

Treatment	Observations
1	2 1 3
2	1 5
3	9 5 6 4
4	3 4 5

Provide a decomposition of the observations and obtain the ANOVA table.

14.6 An ecologist, studying the southwestern Atlantic salt marshes, concentrates on disturbance areas



where dead vegetation creates a mat that kills further growth of the dominant vegetation. These areas provide a place for colonization by nondominant vegetation. In three different salt marshes, he randomly selects sampling plots and counts the number of disturbance areas. Suppose the summary statistics are,

$$\begin{aligned}
 n_1 &= 10 & \bar{y}_1 &= 5 \\
 (n_1 - 1)s_1^2 &= \sum_{j=1}^{10} (y_{1j} - \bar{y}_1)^2 = 30 \\
 n_2 &= 6 & \bar{y}_2 &= 2 \\
 (n_2 - 1)s_2^2 &= \sum_{j=1}^6 (y_{2j} - \bar{y}_2)^2 = 18 \\
 n_3 &= 9 & \bar{y}_3 &= 7 \\
 (n_3 - 1)s_3^2 &= \sum_{j=1}^9 (y_{3j} - \bar{y}_3)^2 = 25
 \end{aligned}$$

Create the ANOVA table.

14.7 Reading levels vary between different magazines. To avoid difficulties caused by different typefaces and sizes, an investigator just counted the number of letters and punctuation signs. Random samples of 20 sentences were selected from *The New Yorker*, *Sports Illustrated*, and *National Geographic*. The resulting summary statistics are:

<i>The New Yorker</i>	<i>Sports Illustrated</i>	<i>National Geographic</i>
$\bar{y}_1 = 94.4$	$\bar{y}_2 = 92.9$	$\bar{y}_3 = 75.5$
$s_1 = 58.4$	$s_2 = 54.2$	$s_3 = 38.1$
$n_1 = 20$	$n_2 = 20$	$n_3 = 20$
where $s_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 / (n_i - 1)$		

Present the ANOVA table for these data.

### 3. POPULATION MODEL AND INFERENCES FOR A COMPLETELY RANDOMIZED DESIGN

To implement a formal statistical test for no difference among treatment effects, we need to have a population model for the experiment. To this end, we assume that the response measurements with the  $i$ th treatment constitute a random sample from a normal population with a mean of  $\mu_i$  and a **common variance** of  $\sigma^2$ . The samples are assumed to be mutually independent.

#### Population Model for Comparing $k$ Treatments

$$Y_{ij} = \mu_i + e_{ij} \quad j = 1, \dots, n_i \quad \text{and} \quad i = 1, \dots, k$$

where  $\mu_i = i$ th treatment mean. The errors  $e_{ij}$  are all independently distributed as  $N(0, \sigma)$ .

Before presenting the test for equality of means based on this model, we give a second parametrization that previews the formulation of statistical models for more complicated designs. For each  $i$ , the mean  $\mu_i$  is considered to be the sum of an overall effect  $\mu$ , common to all treatments, and an effect due only to the  $i$ th treatment.

$$\mu_i = \mu + \alpha_i \quad \text{subject to the constraint} \quad \sum_{i=1}^k n_i \alpha_i = 0$$

The right-hand side is estimated by  $\bar{y} + (\bar{y}_i - \bar{y})$  in the decomposition on page 547 and the estimated treatment effects satisfy the same constraint.

**F DISTRIBUTION**

The  $F$  test will determine if significant differences exist between the  $k$  sample means. The null hypothesis that no difference exists among the  $k$  population means can now be phrased as follows:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

The alternative hypothesis is that not all the  $\mu_i$ 's are equal. Seeking a criterion to test the null hypothesis, we observe that when the population means are all equal,  $\bar{y}_i - \bar{y}$  is expected to be small, and consequently, the treatment mean square  $\sum n_i (\bar{y}_i - \bar{y})^2 / (k - 1)$  is expected to be small. On the other hand, it is likely to be large when the means differ markedly. The error mean square, which provides an estimate of  $\sigma^2$ , can be used as a yardstick for determining how large a treatment mean square should be before it indicates significant differences. Statistical distribution theory tells us that under  $H_0$  the ratio

$$F = \frac{\text{Treatment mean square}}{\text{Error mean square}} = \frac{\text{Treatment SS} / (k - 1)}{\text{Error SS} / \left( \sum_{i=1}^k n_i - k \right)}$$

has an **F distribution** with d.f. =  $(k - 1, n - k)$ , where  $n = \sum n_i$ .

Notice that an  $F$  distribution is specified in terms of its numerator degrees of freedom  $\nu_1 = k - 1$  and denominator degrees of freedom  $\nu_2 = n - k$ . We denote

$$F_\alpha(\nu_1, \nu_2) = \text{Upper } \alpha \text{ point of the } F \text{ distribution with } (\nu_1, \nu_2) \text{ d.f.}$$

which is also called the upper  $100\alpha$ -th percentage point.

The upper  $\alpha = .05$  and  $\alpha = .10$  points are given in Appendix B, Table 6, for several pairs of d.f. With  $\nu_1 = 7$  and  $\nu_2 = 15$ , for  $\alpha = .05$ , we read from column  $\nu_1 = 7$  and row  $\nu_2 = 15$  to obtain  $F_{.05}(7, 15) = 2.71$  (see Table 5).

**TABLE 5** Percentage Points of  $F(\nu_1, \nu_2)$  Distributions  $\alpha = .05$

$\nu_1 \backslash \nu_2$	...	7	...
...			
15	...	2.71	
...			

We summarize the  $F$  test introduced above.

### **$F$ Test for Equality of Means**

Reject  $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$  if

$$F = \frac{\text{Treatment SS}/(k - 1)}{\text{Error SS}/(n - k)} \geq F_\alpha(k - 1, n - k)$$

where  $n = \sum_{i=1}^k n_i$  and  $F_\alpha(k - 1, n - k)$  is the upper  $\alpha$  point of the  $F$  distribution with d.f. =  $(k - 1, n - k)$ .

The computed value of the  $F$ -ratio is usually presented in the last column of the ANOVA table.

### **Example 3** The $F$ Test for Testing the Null Hypothesis of No Difference in Sound Distortion Means

Construct the ANOVA table for the data given in Example 1 concerning a comparison of four brands of earbuds. Test the null hypothesis that the means are equal. Use  $\alpha = .05$ .

**SOLUTION** Using our earlier calculations for the component sums of squares, we construct the ANOVA table that appears in Table 6.

**TABLE 6** ANOVA Table for the Data Given in Example 1

Source	Sum of Squares	d.f.	Mean Square	$F$ -ratio
Treatment	68	3	22.67	$\frac{22.67}{5.22} = 4.34$
Error	94	18	5.22	
Total	162	21		

A test of the null hypothesis  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$  is performed by comparing the observed  $F$  value 4.34 with the tabulated value of  $F$  with d.f. =  $(3, 18)$ . At a .05 level of significance, the tabulated value is found to be 3.16. Because this is exceeded by the observed value, we conclude that there is a significant difference among the four mean sound distortions.

Table 7 gives some typical output from a computer program where the term factor is used instead of treatment. The MINITAB commands for obtaining these results are given in Exercise 14.35.

**TABLE 7** Computer Output: One-Way Analysis of Variance for Sound Distortion Data

One-way Analysis of Variance					
Analysis of Variance					
Source	DF	SS	MS	F	P
Factor	3	68.00	22.67	4.34	0.018
Error	18	94.00	5.22		
Total	21	162.00			

### Exercises

14.8 Using the table of percentage points for the  $F$  distribution, find

(a) The upper 5% point when  $\nu_1 = 7$  and  $\nu_2 = 10$ .

(b) The upper 5% point when  $\nu_1 = 10$  and  $\nu_2 = 7$ .

14.9 Using Appendix B, Table 6, find the upper 10% point of  $F$  for

(a) d.f. = (3, 5)    (b) d.f. = (3, 10)

(c) d.f. = (3, 15)    (d) d.f. = (3, 30)

(e) What effect does increasing the denominator d.f. have?

14.10 A psychologist investigating the connection between music and memory, randomly assigns subjects to one of three treatment groups. The first group hears white noise, the second Mozart, and the third heavy metal. With the appropriate background sound turned on, subjects visually study a picture. The picture is removed, the sound turned off, and then the subjects are asked to answer a questionnaire concerning the image. Suppose the number of errors result in the following ANOVA table,

Source	Sum of Squares	d.f.
Treatment	12	2
Error	104	41
Total	116	43

Carry out the  $F$  test for equality of means taking  $\alpha = .10$ .

14.11 Based on the current General Social Survey, an index is created from the few questions asking about the degree of confidence in government. Six age groups are compared. Suppose the values of the index result in the following ANOVA table.

Source	Sum of Squares	d.f.
Treatment	23	5
Error	56	30
Total	79	35

Carry out the  $F$  test for equality of means taking  $\alpha = .05$ .

14.12 Using the data from Exercise 14.1, test for equality of means using  $\alpha = .05$ .

14.13 Test for equality of means based on the data in Exercise 14.2. Take  $\alpha = .05$ .

14.14 Three bread recipes are to be compared with respect to density of the loaf. Five loaves will be baked using each recipe.

(a) If one loaf is made and baked at a time, how would you select the order?

(b) Given the following data, conduct an  $F$  test for equality of means. Take  $\alpha = .05$ .

Recipe	Observation				
1	.95	.86	.71	.72	.74
2	.71	.85	.62	.72	.64
3	.69	.68	.51	.73	.44

- 14.15 Test for equality of means based on the data in Exercise 14.3. Take  $\alpha = .05$ .
- 14.16 Refer to the data on reading levels in Exercise 14.7. Test for equality of means. Take  $\alpha = .05$ .

## 4. SIMULTANEOUS CONFIDENCE INTERVALS

The ANOVA  $F$  test is only the initial step in our analysis. It determines if significant differences exist among the treatment means. Our goal should be more than to merely conclude that treatment differences are indicated by the data. Rather, we must detect likenesses and differences among the treatments. Thus, the problem of estimating differences in treatment means is of even greater importance than the overall  $F$  test.

Referring to the comparison of  $k$  treatments using the data structure given in Table 1, let us examine how a confidence interval can be established for  $\mu_1 - \mu_2$ , the mean difference between treatment 1 and treatment 2. The statistic

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\text{SSE}}{n - k} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

has a  $t$  distribution with d.f. =  $n - k$ , and this can be employed to construct a confidence interval for  $\mu_1 - \mu_2$ . More generally:

### Confidence Interval for a Single Difference

A  $100(1 - \alpha)\%$  confidence interval for  $\mu_i - \mu_{i'}$ , the difference of means for treatment  $i$  and treatment  $i'$  is given by

$$(\bar{Y}_i - \bar{Y}_{i'}) \pm t_{\alpha/2} S \sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}$$

where

$$S = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n - k}}$$

and  $t_{\alpha/2}$  is the upper  $\alpha/2$  point of  $t$  with d.f. =  $n - k$ .

If the  $F$  test first shows a significant difference in means, then some statisticians feel that it is reasonable to compare means pairwise according to the preceding intervals. However, many statisticians prefer a more conservative procedure based on the following reasoning.

Without the provision that the  $F$  test is significant, the preceding method provides **individual** confidence intervals for pairwise differences. However, with  $k = 4$  treatments, there are  $\binom{4}{2} = 6$  pairwise differences  $\mu_i - \mu_{i'}$ , and this procedure applied to all pairs yields six confidence statements, each having a  $100(1 - \alpha)\%$  level of confidence. It is difficult to determine what level of confidence will be achieved for claiming that **all** six of these statements are correct. To overcome this dilemma, procedures have been developed for several confidence intervals to be constructed in such a manner that the joint probability that all the statements are true is guaranteed not to fall below a predetermined level. Such intervals are called **multiple confidence intervals** or **simultaneous confidence intervals**. Numerous methods proposed in the statistical literature have achieved varying degrees of success. We present one that can be used simply and conveniently in general applications.

The procedure, called the **multiple- $t$  confidence intervals**, consists of setting confidence intervals for the differences  $\mu_i - \mu_{i'}$  in much the same way we just did for the individual differences, except that a different percentage point is read from the  $t$  table.

Operationally, the construction of these confidence intervals does not require any new concepts or calculations, but it usually involves some nonstandard percentage point of  $t$ . For example, with  $k = 3$  and  $1 - \alpha = .95$ , if we want to set simultaneous intervals for all  $m = \binom{k}{2} = 3$  pairwise differences, we require that the upper  $\alpha/(2m) = .05/6 = .00833$  point of a  $t$  distribution.

### Multiple- $t$ Confidence Intervals

A set of  $100(1 - \alpha)\%$  simultaneous confidence intervals for  $m$  number of pairwise differences  $\mu_i - \mu_{i'}$  is given by

$$(\bar{Y}_i - \bar{Y}_{i'}) \pm t_{\alpha/2m} S \sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}$$

where  $S = \sqrt{\text{MSE}}$ ,  $m =$  the number of confidence statements, and  $t_{\alpha/2m} =$  the upper  $\alpha/(2m)$  point of  $t$  with d.f.  $= n - k$ .

Prior to sampling, the probability of all the  $m$  statements being correct is at least  $1 - \alpha$ .

#### Example 4 Calculating Multiple- $t$ Confidence Intervals to Reveal Which Means Differ

An experiment is conducted to determine the soil moisture deficit resulting from varying amounts of residual timber left after cutting trees in a forest. The three treatments are treatment 1: no timber left; treatment 2: 2000 bd ft left; treatment 3: 8000 bd ft left. (Board feet is a particular unit of measurement of

timber volume.) The measurements of moisture deficit are given in Table 8. Perform the ANOVA test and construct 95% multiple- $t$  confidence intervals for the treatment differences.

**SOLUTION** Our analysis employs convenient alternative forms of the expressions for sums of squares involving totals.

**TABLE 8** Moisture Deficit in Soil

Treatment	Observations	Total	Mean
1	1.52 1.38 1.29 1.48 1.63	$T_1 = 7.30$	$\bar{y}_1 = 1.460$
2	1.63 1.82 1.35 1.03 2.30 1.45	$T_2 = 9.58$	$\bar{y}_2 = 1.597$
3	2.56 3.32 2.76 2.63 2.12 2.78	$T_3 = 16.17$	$\bar{y}_3 = 2.695$
		Grand total $T = 33.05$	Grand mean $\bar{y} = 1.944$

The total number of observations  $n = 5 + 6 + 6 = 17$ .

$$\begin{aligned} \text{Total SS} &= \sum_{i=1}^3 \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^3 \sum_{j=1}^{n_i} y_{ij}^2 - \frac{T^2}{n} \\ &= 71.3047 - 64.2531 = 7.0516 \end{aligned}$$

$$\begin{aligned} \text{Treatment SS} &= \sum_{i=1}^3 n_i (\bar{y}_i - \bar{y})^2 = \sum_{i=1}^3 \frac{T_i^2}{n_i} - \frac{T^2}{n} \\ &= 69.5322 - 64.2531 = 5.2791 \end{aligned}$$

$$\text{Error SS} = \text{Total SS} - \text{Treatment SS} = 1.7725$$

The ANOVA table appears in Table 9.

**TABLE 9** ANOVA Table for Comparison of Moisture Deficit

Source	Sum of Squares	d.f.	Mean Square	$F$ -ratio
Treatment	5.2791	2	2.640	20.8
Error	1.7725	14	.127	
Total	7.0516	16		

Because the observed value of  $F$  is larger than the tabulated value  $F_{.05}(2, 14) = 3.74$ , the null hypothesis of no difference in the treatment effects is rejected at  $\alpha = .05$ . In fact, this would be true at almost any significance level. In constructing a set of 95% multiple- $t$  confidence intervals for pairwise differences, note that there are  $\binom{3}{2} = 3$  pairs, so

$$\frac{\alpha}{2m} = \frac{.05}{(2 \times 3)} = .00833$$

From Appendix B, Table 4, the upper .00833 point of  $t$  with d.f. = 14 is 2.718. The simultaneous confidence intervals are calculated as follows:

$$\begin{aligned}\mu_2 - \mu_1: & (1.597 - 1.460) \pm 2.718 \times .356 \times \sqrt{\frac{1}{6} + \frac{1}{5}} \\ & = (-.45, .72)\end{aligned}$$

$$\begin{aligned}\mu_3 - \mu_2: & (2.695 - 1.597) \pm 2.718 \times .356 \times \sqrt{\frac{1}{6} + \frac{1}{6}} \\ & = (.54, 1.66)\end{aligned}$$

$$\begin{aligned}\mu_3 - \mu_1: & (2.695 - 1.460) \pm 2.718 \times .356 \times \sqrt{\frac{1}{6} + \frac{1}{5}} \\ & = (.65, 1.82)\end{aligned}$$

These confidence intervals indicate that treatments 1 and 2 do not differ appreciably, but the mean for treatment 3 is considerably higher than the means for treatments 1 and 2.

## Exercises

- 14.17 Taking  $\alpha = .05$  and  $n - k = 26$ , determine the appropriate percentile of the  $t$  distribution when calculating the multiple- $t$  confidence intervals with (a)  $m = 3$  and (b)  $m = 5$ .
- 14.18 Construct the 90% multiple- $t$  confidence intervals using the sound distortion data in Example 1.
- 14.19 As suggested in the photo on the front piece of this chapter, consumers can rate new HDTVs. Many Web sites use a five point scale. Although the individual responses are not normal, the central limit result applied to each of the four samples does justify treating the treatment means as if the normal assumption prevailed. The summary statistics of ratings for four of the most popular sets, at one Web site are
- |            |                    |            |
|------------|--------------------|------------|
| $n_1 = 14$ | $\bar{y}_1 = 4.50$ |            |
| $n_2 = 20$ | $\bar{y}_2 = 3.60$ | $s = 1.56$ |
| $n_3 = 30$ | $\bar{y}_3 = 4.13$ |            |
| $n_4 = 15$ | $\bar{y}_4 = 4.56$ |            |
- Use  $\alpha = .10$  and determine:
- (a)  $t$  intervals for each of the six differences of means.
- (b) The six multiple- $t$  intervals.
- 14.20 Refer to the data on reading levels in Exercise 14.7.
- (a) Calculate simultaneous confidence intervals for the differences in means.
- (b) Are you surprised by the conclusion regarding reading levels? Give another variable that might better quantify reading levels.
- 14.21 Determine the expression for the length of the  $t$  interval for  $\mu_1 - \mu_2$  and the multiple- $t$  interval for  $\mu_1 - \mu_2$  when  $m = 10$ . The ratio of lengths does not depend on the data. Evaluate this ratio for  $\alpha = .10$  and  $n - k = 15$ .



## 5. GRAPHICAL DIAGNOSTICS AND DISPLAYS TO SUPPLEMENT ANOVA

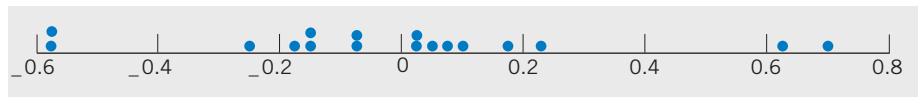
In addition to testing hypotheses and setting confidence intervals, an analysis of data must include a critical examination of the assumptions involved in a model. As in regression analysis, of which analysis of variance is a special case, the residuals must be examined for evidence of serious violations of the assumptions. This aspect of the analysis is ignored in the ANOVA table summary.

### Example 5 Plotting Residuals

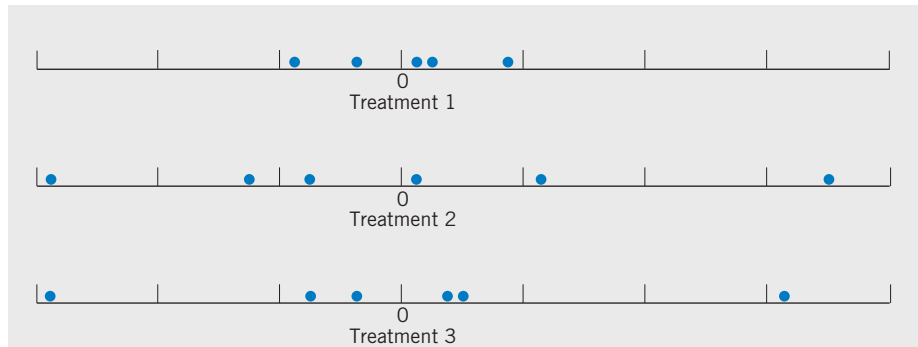
Determine the residuals for the moisture data given in Example 4 (see Table 10) and graphically examine them for possible violations of the assumptions.

**SOLUTION** **TABLE 10** Residuals  $y_{ij} - \bar{y}_i$  for the Data Given In Table 8

Treatment	Residuals					
1	.06	-.08	-.17	.02	.17	
2	.03	.22	-.25	-.57	.70	-.15
3	-.14	.63	.07	-.07	-.58	.09



(a) Combined residual plot



(b) Residuals with individual treatment

Figure 1 Residual plots for the data given in Example 4.

The residual plots of these data are shown in Figure 1, where the combined dot diagram is presented in (a) and the dot diagrams of residuals corresponding to individual treatments appear in (b).

From an examination of the dot diagrams, the variability in the points for treatment 1 appears to be somewhat smaller than the variabilities in the points for treatments 2 and 3. However, given so few observations, it is difficult to determine if this has occurred by chance or if treatment 1 actually has a smaller variance. A few more observations are usually necessary to obtain a meaningful pattern for the individual treatment plots.

Fortunately, the ANOVA testing procedure is **robust** in the sense that small or moderate departures from normality and constant variance do not seriously affect its performance.

In addition to the ANOVA a graphical portrayal of the data, as a box plot for each treatment, conveys important information available for making comparisons of populations.

### Example 6 Box Plots Reveal Differences between Populations

The sepal width was measured on 50 iris flowers for each of three varieties, *Iris setosa*, *Iris versicolor*, and *Iris virginica*. A computer calculation produced the summary shown in the following ANOVA table.

Source	SS	d.f.	<i>F</i>
Treatment	11.345	2	49.16
Error	16.962	147	
Total	28.307	149	

Since  $F_{.05}(2, 147) \approx 3.05$ , we reject the null hypothesis of equal sepal width means at the 5% level of significance.

Treatment	Sample Mean
<i>Iris setosa</i>	3.428
<i>Iris versicolor</i>	2.770
<i>Iris virginica</i>	2.974

A calculation of multiple-*t* confidence intervals shows that all population means differ from one another (see Exercise 14.36). Display the data on the three varieties in boxplots.

**SOLUTION** The data are given in Exercise 14.36. Boxplots graphically display the variation in the sepal width measurements. From Figure 2, we see that the *Iris setosa* typically has larger sepal width. R. A. Fisher, who developed analysis of variance, used these data along with other lengths and widths to introduce a statistical technique for identifying varieties of plants.

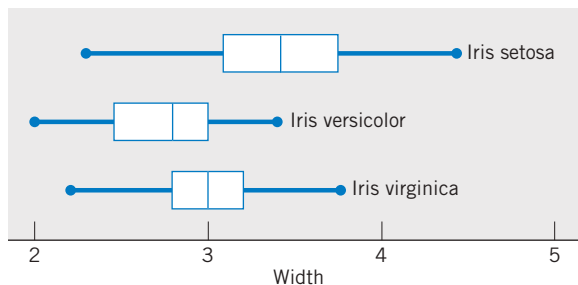


Figure 2 Boxplots for the three iris samples.

## 6. RANDOMIZED BLOCK EXPERIMENTS FOR COMPARING $k$ TREATMENTS

Just as we can pair like subjects or experimental units to improve upon the procedure of taking two independent samples, we can also arrange, or **block** subjects into homogeneous groups of size  $k$  when comparing  $k$  treatments. Then if each treatment is applied to exactly one unit in the block and comparisons are only drawn between treatment responses from the same block, extraneous variability should be greatly reduced. It is this concept of **blocking** that underlies the **randomized block design**.

The term “block design” originated from the design of agricultural field trials, where “block” refers to a group of adjacent plots. A few typical examples for which the block design may be appropriate are clinical trials to compare several competing drugs, where the experimental subjects are grouped in blocks according to age group and severity of symptoms; psychological experiments comparing several stimuli, where subjects may be blocked according to socioeconomic background; and comparison of several techniques for storing fruit or vegetables, where each incoming shipment is regarded as a block.

As its name implies, **randomization** is a basic part of the block design. This time, once the grouping of like experimental subjects in blocks is accomplished, we randomly select one subject from the first block to receive treatment 1, one of the remaining subjects to receive treatment 2, and so on. The same procedure is repeated with a new randomization for each of the remaining blocks.

Once the data are obtained, they can be arranged in rows according to the treatments and in columns according to the blocks. If we designate the measurement corresponding to treatment  $i$  and block  $j$  by  $y_{ij}$ , the data structure of a randomized block design with  $b$  blocks and  $k$  treatments is shown in Table 11. The

**TABLE 11** Data Structure of a Randomized Block Design with  $b$  Blocks and  $k$  Treatments

	Block 1	Block 2	...	Block $b$	Treatment Means
Treatment 1	$y_{11}$	$y_{12}$	...	$y_{1b}$	$\bar{y}_{1\cdot}$
Treatment 2	$y_{21}$	$y_{22}$	...	$y_{2b}$	$\bar{y}_{2\cdot}$
.	.	.	...	.	.
.	.	.	...	.	.
.	.	.	...	.	.
Treatment $k$	$y_{k1}$	$y_{k2}$	...	$y_{kb}$	$\bar{y}_{k\cdot}$
Block means	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	...	$\bar{y}_{\cdot b}$	$\bar{y}_{\cdot\cdot}$

row and column means are denoted by

$$i\text{th treatment (row) mean } \bar{y}_{i\cdot} = \frac{1}{b} \sum_{j=1}^b y_{ij}$$

$$j\text{th block (column) mean } \bar{y}_{\cdot j} = \frac{1}{k} \sum_{i=1}^k y_{ij}$$

and the overall mean, or

$$\text{grand mean } \bar{y}_{\cdot\cdot} = \frac{1}{bk} \sum_{i=1}^k \sum_{j=1}^b y_{ij}$$

These means are shown in the margins of the table. Here an overbar on  $y$  indicates an average and a dot in the subscript denotes that the average is taken over the subscript appearing in that place.

We now discuss the analysis of variance for a randomized block design with illustrative calculations based on the data given in Example 7.

**Example 7** The Structure of Data from a Randomized Block Experiment

The cutting speeds of four types of tools are being compared in an experiment. Five materials of varying degrees of hardness are to be used as experimental blocks. The data pertaining to measurements of cutting time in seconds appear in Table 12.

**TABLE 12** Measurements of Cutting Time According to Types of Tool (Treatments) and Hardness of Material (Blocks)

Treatment	Block					Treatment Means
	1	2	3	4	5	
1	12	2	8	1	7	$\bar{y}_{1.} = 6$
2	20	14	17	12	17	$\bar{y}_{2.} = 16$
3	13	7	13	8	14	$\bar{y}_{3.} = 11$
4	11	5	10	3	6	$\bar{y}_{4.} = 7$
Block means	$\bar{y}_{.1} = 14$	$\bar{y}_{.2} = 7$	$\bar{y}_{.3} = 12$	$\bar{y}_{.4} = 6$	$\bar{y}_{.5} = 11$	$\bar{y}_{..} = \frac{200}{20} = 10$

The observations form a two-way table, and their decomposition indicates both a term for row (treatment) deviations and column (block) deviations.

**Decomposition of Observations**

Observation = Grand mean + Deviation due to treatment + Deviation due to block + Residual

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$

**TABLE 13** Decomposition of Observations for the Randomized Block Experiment in Table 12

Observation $y_{ij}$	=	Grand Mean $\bar{y}_{..}$	+	Treatment Effect $\bar{y}_{i.} - \bar{y}_{..}$
$\begin{bmatrix} 12 & 2 & 8 & 1 & 7 \\ 20 & 14 & 17 & 12 & 17 \\ 13 & 7 & 13 & 8 & 14 \\ 11 & 5 & 10 & 3 & 6 \end{bmatrix}$	=	$\begin{bmatrix} 10 & 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 & 10 \end{bmatrix}$	+	$\begin{bmatrix} -4 & -4 & -4 & -4 & -4 \\ 6 & 6 & 6 & 6 & 6 \\ 1 & 1 & 1 & 1 & 1 \\ -3 & -3 & -3 & -3 & -3 \end{bmatrix}$
		+		Residual $y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$
		Block Effect $\bar{y}_{.j} - \bar{y}_{..}$	+	
		$\begin{bmatrix} 4 & -3 & 2 & -4 & 1 \\ 4 & -3 & 2 & -4 & 1 \\ 4 & -3 & 2 & -4 & 1 \\ 4 & -3 & 2 & -4 & 1 \end{bmatrix}$	+	$\begin{bmatrix} 2 & -1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ -2 & -1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 0 & -2 \end{bmatrix}$

Taking the observation  $y_{11} = 12$  in Example 7, we obtain

$$\begin{aligned} 12 &= 10 + (6 - 10) + (14 - 10) + (12 - 14 - 6 + 10) \\ &= 10 + (-4) + (4) + (2) \end{aligned}$$

Table 13 contains the results of the decomposition of all the observations. Table 14 gives an alternative format.

**TABLE 14** An Alternative Format of the Decomposition Table for a Randomized Block Experiment

Treatment	Block					Deviation of Treatment Mean from Grand Mean
	1	...	$j$	...	$b$	
1						$(\bar{y}_{1.} - \bar{y}_{..})$
.			.			.
.			.			.
.			.			.
$i$		$\cdots(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})\cdots$				$(\bar{y}_{i.} - \bar{y}_{..})$
.			.			.
.			.			.
.			.			.
$k$						$(\bar{y}_{k.} - \bar{y}_{..})$
Deviation of block mean from grand mean	$(\bar{y}_{.1} - \bar{y}_{..})$	$\cdots$	$(\bar{y}_{.j} - \bar{y}_{..})$	$\cdots$	$(\bar{y}_{.b} - \bar{y}_{..})$	Grand mean $\bar{y}_{..}$

The number of distinct entries  $\bar{y}_{i.} - \bar{y}_{..}$  in the treatment effects array is  $k$ , and the single constraint is that they must sum to zero. Thus,  $k - 1$  degrees of freedom are associated with the treatment sum of squares.

**Sum of Squares Due to Treatment**

$$\begin{aligned} SS_T &= \sum_{i=1}^k \sum_{j=1}^b (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &= b \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 \end{aligned}$$

with d.f. =  $k - 1$ .

In the array of treatment effects in Table 13, each entry appears  $b = 5$  times, once in each block. The treatment sum of squares for this example is then

$$SS_T = 5(-4)^2 + 5(6)^2 + 5(1)^2 + 5(-3)^2 = 310$$

with d.f. =  $4 - 1 = 3$ .

In a similar manner, we obtain the **block sum of squares**.

### Sum of Squares Due to Block

$$\begin{aligned} SS_B &= \sum_{i=1}^k \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 \\ &= k \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 \end{aligned}$$

with d.f. =  $b - 1$ .

Referring to the array of block effects in Table 13, we find the block sum of squares for our example to be

$$SS_B = 4(4)^2 + 4(-3)^2 + 4(2)^2 + 4(-4)^2 + 4(1)^2 = 184$$

with d.f. =  $5 - 1 = 4$ .

The number of degrees of freedom associated with the residual array is  $(b - 1)(k - 1)$ . To understand why this is so, note that among the  $b \times k$  residuals, the following constraints are satisfied. One constraint is that the sum of all entries must be zero. The fact that all row sums are zero introduces  $k - 1$  additional constraints. This is so because having fixed any  $k - 1$  row totals and the grand total, the remaining row total is automatically fixed. By the same reasoning,  $b - 1$  additional constraints arise from the fact that all column totals are zero. Consequently, the number of degrees of freedom for the residual is

$$bk - 1 - (k - 1) - (b - 1) = (b - 1)(k - 1)$$

### Residual Sum of Squares or Error Sum of Squares

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

with d.f. =  $(b - 1)(k - 1)$ .

In our example,

$$SSE = 2^2 + 0^2 + (-2)^2 + 0^2 + \dots + (-2)^2 = 24$$

with d.f. = (5 - 1)(4 - 1) = 12.

Finally, the total sum of squares equals the sum of squares of each observation about the grand mean, or

$$\text{Total sum of squares} = \sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^k \sum_{j=1}^b y_{ij}^2 - bk\bar{y}_{..}^2.$$

with d.f. =  $bk - 1$ . In our example, we sum the square of each entry in the array for  $y_{ij}$  and subtract the sum of squares of entries in the  $\bar{y}_{..}$  array.

$$\begin{aligned} \text{Total sum of squares} &= (12)^2 + (2)^2 + \dots + (6)^2 \\ &\quad - [(10)^2 + (10)^2 + \dots + (10)^2] \\ &= 518 \end{aligned}$$

With d.f. = (5)(4) - 1 = 19. This provides a check on our previous calculations, because those sums of squares and degrees of freedom must sum to these totals.

These calculations are conveniently summarized in the ANOVA tables shown in Table 15 for the general case and Table 16 for the data in our numerical example. The last column of  $F$ -ratios will be explained after we discuss the population model.

**TABLE 15** ANOVA Table for a Randomized Block Design

Source	Sum of Squares	d.f.	Mean Square	$F$ -Ratio
Treatments	$SS_T = b \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2$	$k - 1$	$MS_T = \frac{SS_T}{k - 1}$	$\frac{MS_T}{MSE}$
Blocks	$SS_B = k \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2$	$b - 1$	$MS_B = \frac{SS_B}{b - 1}$	$\frac{MS_B}{MSE}$
Residual	$SSE = \sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$(b - 1)(k - 1)$	$MSE = \frac{SSE}{(b - 1)(k - 1)}$	
Total	$\sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2$	$bk - 1$		

**TABLE 16** ANOVA Table for the Data Given in Example 7

Source	Sum of Squares	d.f.	Mean Square	$F$ -Ratio
Treatments	310	3	103.3	51.7
Blocks	184	4	46	23
Residual	24	12	2	
Total	518	19		



Again, a statistical test of treatment differences is based on an underlying population model.

### Population Model for a Randomized Block Experiment

$$\begin{aligned} \text{Observation} &= \text{Overall mean} + \text{Treatment effect} + \text{Block effect} + \text{Error} \\ Y_{ij} &= \mu + \alpha_i + \beta_j + e_{ij} \end{aligned}$$

for  $i = 1, \dots, k$  and  $j = 1, \dots, b$ , where the parameters satisfy

$$\sum_{i=1}^k \alpha_i = 0 \quad \sum_{j=1}^b \beta_j = 0$$

and the  $e_{ij}$  are random errors independently distributed as  $N(0, \sigma)$ .

Tests for the absence of treatment differences or differences in block effects can now be performed by comparing the corresponding mean square with the yardstick of the error mean square by using an  $F$  test.

Reject  $H_0: \alpha_1 = \dots = \alpha_k = 0$  (no treatment differences) if

$$\frac{MS_T}{MSE} > F_\alpha(k-1, (b-1)(k-1))$$

Reject  $H_0: \beta_1 = \dots = \beta_b = 0$  (no block differences) if

$$\frac{MS_B}{MSE} > F_\alpha(b-1, (b-1)(k-1))$$

To test the hypothesis of no treatment differences for the analysis of variance in Table 16, we find that the tabulated .05 point of  $F(3, 12)$  is 3.49, a value far exceeded by the observed  $F$ -ratio for treatment effect. We therefore conclude that a highly significant treatment difference is indicated by the data. The block effects are also highly significant, because the observed  $F$  value of 23 is much larger than the tabulated value  $F_{.05}(4, 12) = 3.26$ .

Again, we stress that a serious violation of the model assumptions is likely to jeopardize the conclusions drawn from the preceding analyses and a careful examination of the residuals should be an integral part of the analysis. In addition to plotting the whole set of residuals in a graph, separate plots for individual treatments and individual blocks should also be studied. When observations are collected over time, a plot of the residuals versus the time order is also important.

### CONFIDENCE INTERVALS FOR TREATMENT DIFFERENCES

In addition to performing the overall  $F$  test for detecting treatment differences, the experimenter typically establishes confidence intervals to compare specific pairs of treatments. This is particularly important when the  $F$  test leads to a rejection of the null hypothesis, thus signifying the presence of treatment differences.

We next estimate the difference  $\alpha_i - \alpha_{i'}$  of the mean responses of treatments  $i$  and  $i'$ . Because  $\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}$  is normally distributed with a mean of  $\alpha_i - \alpha_{i'}$  and a variance of

$$\sigma^2 \left( \frac{1}{b} + \frac{1}{b} \right) = \sigma^2 \left( \frac{2}{b} \right)$$

the ratio

$$T = \frac{(\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}) - (\alpha_i - \alpha_{i'})}{\sqrt{\text{MSE}} \sqrt{2/b}}$$

has a  $t$  distribution with d.f. =  $(b - 1)(k - 1)$ . This result can be used to construct a confidence interval for an individual difference  $\alpha_i - \alpha_{i'}$ .

When several such pairwise comparisons are to be integrated into a combined confidence statement, the concept of simultaneous confidence intervals, discussed in Section 4, is again applied.

### Exercises

- 14.22 Suppose you wish to compare three different brands of tick collars for dogs. You have available three of each of the breeds Poodle, Lab, Collie, and Dachshund. Explain how you would assign a brand of tick collar to each of the 12 dogs in order to conduct a randomized block experiment.
- 14.23 The hours of relief are measured under a placebo, Brand A, and Brand B sinus medicines. Call these treatment 1, treatment 2, and treatment 3, respectively. They are each given in succession, in random order, to a subject with two days between each treatment. Suppose the data are
- 14.24 The yield, in pounds, of three types of heritage tomatoes is obtained after planting them in three equal sized plots within a site. The assignment of type to a plot is random. A total of four different sites are used. Suppose the data are

Treatment	Block			
	1	2	3	4
1	11	10	7	0
2	7	8	7	2
3	15	6	13	10

Treatment	Block			
	1	2	3	4
1	35	24	28	21
2	19	14	14	13
3	21	16	21	14

- (a) Provide a decomposition for the observations from this randomized block experiment.

- (b) Find the sum of squares for each array.
  - (c) Determine the degrees of freedom by checking the constraints for each array.
- 14.25 Refer to Exercise 14.23. Present the ANOVA table. What conclusions can you draw from the two  $F$  tests? Take  $\alpha = .05$ .
- 14.26 Refer to Exercise 14.24. Present the ANOVA table. What conclusions can you draw from the two  $F$  tests? Take  $\alpha = .05$ .
- 14.27 Three loaves of bread, each made according to a different recipe, are baked in one oven at the same time. Because of possible uncontrolled variations in oven performance, each baking is treated as a block. This procedure is repeated five times, and the following measurements of density are obtained.

Recipe	Block				
	1	2	3	4	5
1	.95	.86	.71	.72	.74
2	.71	.85	.62	.72	.64
3	.69	.88	.51	.73	.44

- (a) How should the three oven positions of the three loaves be selected for each trial?

- (b) Perform an analysis of variance for these data.

14.28 Referring to Exercise 14.27:

- (a) Obtain simultaneous confidence intervals for the pairwise differences in mean density for the three recipes. Take  $\alpha = .05$ .
- (b) Calculate the residuals and make a normal-scores plot.

14.29 As part of a cooperative study on the nutritional quality of oats, 6 varieties of oat kernels with their hulls removed are subjected to a mineral analysis. The plants are grown according to a randomized block design, and the measurements of protein by percent of dry weight are recorded in Table 17.

- (a) Perform an analysis of variance for these data.
- (b) Calculate and plot the residuals. Does the model appear to be adequate?

14.30 Referring to Exercise 14.29, suppose that variety 6 is of special interest. Construct simultaneous 90% confidence intervals for the differences between the mean of variety 6 and each of the other means.

**TABLE 17** Protein (prct. dry wt.)

Treatment	Block					
	1	2	3	4	5	6
1	19.09	20.29	20.31	19.60	18.62	20.10
2	16.28	17.88	16.88	17.57	16.72	17.32
3	16.31	18.17	17.38	17.53	16.34	17.88
4	17.50	18.05	17.59	17.64	17.38	18.04
5	16.25	16.92	15.88	14.78	15.97	16.66
6	21.09	21.37	21.38	20.52	21.09	21.58

Data courtesy of D. Peterson, L. Schrader, and V. Youngs.

**USING STATISTICS WISELY**

1. When collecting data according to a one-way ANOVA design, conduct the trials in random order if at all possible.
2. Do not routinely accept the analysis of variance generated by statistical software. Instead, inspect the residuals for outliers or patterns indicating that the variance is not constant across treatments.

## KEY IDEAS AND FORMULAS

Several populations can be compared using the **Analysis of variance (ANOVA)**.

A **completely randomized design** specifies taking independent random samples from each population.

All of the observations are assumed to have a **common variance**. Then the ANOVA consists of a separation of the total sum of squares into components due to different sources of variation.

When samples are taken from  $k$  populations the two sources of variation are the **within** population variation or **error** and the differences **between** population means or **treatments**. The  $j$ th observation on the  $i$ th treatment is  $y_{ij}$ . The error variation is estimated using the **residuals** or  $y_{ij} - \bar{y}_i$  which are the deviations of the observations from their respective sample means.

The total variation in the observations  $y_{ij}$  is expressed as

$$\text{Total sum of squares} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$  is partitioned into the two components:

1. **Treatment sum of squares:**

$$SS_T = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$$

2. **Error sum of squares:**

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

This decomposition of sums of squares is summarized in an **analysis of variance (ANOVA) table**. This table includes the mean square for each sum of squares.

$$\text{Mean square} = \frac{\text{Sum of squares}}{\text{Degrees of freedom}}$$

The **F statistic** = (Treatment mean square) / (Error mean square)

which has an **F distribution** with  $k - 1$  and  $n - k$  degrees of freedom.

To ensure an overall confidence level for **all** confidence intervals, it is desirable to use a **simultaneous (multiple) confidence intervals** procedure when calculating confidence intervals for the many mean differences. The **multiple-t confidence intervals** for all  $m = k(k - 1)/2$  statements about the mean differences  $\mu_i - \mu_{i'}$  are

$$\left( \bar{Y}_i - \bar{Y}_{i'} - t_{\alpha/2m} \sqrt{\frac{\text{MSE}}{n - k}}, \quad \bar{Y}_i - \bar{Y}_{i'} + t_{\alpha/2m} \sqrt{\frac{\text{MSE}}{n - k}} \right)$$

Fortunately, the one-way analysis of variance is **robust** with respect to small or moderate departures from the assumptions of normal errors with common variance.

The technique of **blocking**, or grouping experimental units into homogeneous sets, can reduce extraneous variation and sharpen comparisons among treatments.

The analysis of variance of a **randomized block experiment** is based on the partition of the total sum of squares  $\sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2$  into three components:

1. **Treatment sum of squares:**

$$SS_T = b \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2$$

2. **Block sum of squares:**

$$SS_B = k \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2$$

3. **Error sum of squares:**

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

## TECHNOLOGY

### *One-way analysis of variance (ANOVA)*

#### **MINITAB**

We illustrate with the data from different populations in separate columns. For example, with data from 3 populations in columns C1, C2, and C3:

Data

```
C1:  6  10  8
C2: 13  14 15 14
C3:  3   7  5
```

**Stat > ANOVA > One-way (Unstacked)**

Type C1–C3 in **Responses (in separate columns):**

Click on **Graphs**. Select **Three in one**. Click **OK**.

Click **OK**.

In addition to the analysis, this produces a histogram of the residuals as well as a plot of residuals versus fitted values.

#### **EXCEL**

Enter the data from the different populations in separate columns, for instance A, B, C, etc., with a label in the first cell of each column.

Select **Tools**, then **Data Analysis**. Select **Statistical** and then **Anova:Single Factor**.

Click **OK**. Highlight the columns and check **Labels in First Row**.

Enter **Alpha** and click **OK**.

The output includes the ANOVA table and  $P$ -value for test of equality of the means.

### TI-84/83 PLUS

Store the data in lists, one for each population. With 3 populations enter data in **L1**, **L2**, and **L3**.

Select **STAT > TESTS > ANOVA**.

Enter **L1**, **L2**, **L3** and close the parenthesis to obtain **ANOVA(L1, L2, L3)**.

Press **Enter**.

The software returns the value of the  $F$  statistic and the  $P$ -value, along with other values from the ANOVA table.

## 7. REVIEW EXERCISES

- 14.31 Food scientists investigated whether making a cheese sauce by a continuous process or batch process made any difference to taste. They also included a third treatment, a carefully selected ideal product, that served as a control. Treatment 1 is the ideal, treatment 2 is the continuous process, and treatment 3 is the batch process. Suppose the sensory ratings by trained testers are

Treatment 1	Treatment 2	Treatment 3
19	16	13
18	11	16
21	13	18
18	14	11
	11	15
		11

Provide a decomposition for the observations above from a completely randomized design with three treatments.

- 14.32 Compute the sums of squares and construct the ANOVA table for the data given in Exercise 14.31.

- 14.33 Using the table of percentage points for the  $F$  distribution, find

- (a) The upper 5% point when d.f. = (7, 13).  
 (b) The upper 5% point when d.f. = (7, 20).  
 (c) The upper 10% point when d.f. = (7, 12).

- 14.34 As part of the multilab study, four fabrics are tested for flammability at the National Bureau of Standards. The following burn times in seconds are recorded after a paper tab is ignited on the hem of a dress made of each fabric.

Fabric 1	Fabric 2	Fabric 3	Fabric 4
17.8	11.2	11.8	14.9
16.2	11.4	11.0	10.8
17.5	15.8	10.0	12.8
17.4	10.0	9.2	10.7
15.0	10.4	9.2	10.7

- (a) State the statistical model and present the ANOVA table. With  $\alpha = .05$ , test the null hypothesis of no difference in the degree of flammability for the four fabrics.

- (b) If the null hypothesis is rejected, construct simultaneous confidence intervals to determine the fabric(s) with the lowest mean burn time.
- (c) Plot the residuals and comment on the plausibility of the assumptions.
- (d) If the tests had been conducted one at a time on a single mannequin, how would you have randomized the fabrics tested in this experiment?

**The Following Exercises Require a Computer**

14.35 *Using the computer.* MINITAB can be used for ANOVA. Start with the data on each treatment, from Example 1, set in separate columns. The sequence of commands and output is:

**Data**

C1: 10 15 8 12 15  
 C2: 14 18 21 15  
 C3: 17 16 14 15 17 15 18  
 C4: 12 15 17 15 16 15

**Dialog box:**

Stat > ANOVA > One-way (Unstacked)  
 Type C1-C4 in Responses. Click OK.

The **Output** is as follows:

**One-way ANOVA: C1, C2, C3, C4**

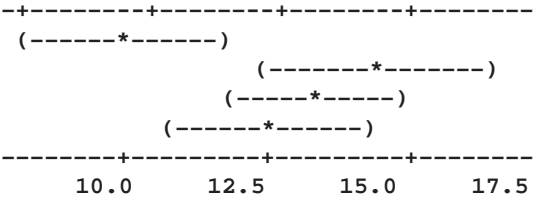
SOURCE	DF	SS	MS	F	P
FACTOR	3	68.00	22.67	4.34	0.018
ERROR	18	94.00	5.22		
TOTAL	21	162.00			

S = 2.285 R-Sq= 41.98%

LEVEL	N	MEAN	STDEV
C1	5	12.000	3.082
C2	4	17.000	3.162
C3	7	16.000	1.414
C4	6	15.000	1.673

POOLED STDEV = 2.285

**INDIVIDUAL 95% CIS FOR MEAN  
 BASED ON POOLED STDEV**



Use computer software to analyze the moisture data in Table 8.

14.36 The iris data described in Example 6 are given in the stem-and-leaf diagrams below.

LEAF UNIT = 0.10

SETOSA	VERSICOLOR	VIRGINICA
	2 0	
2 3	2 22333	2 2
2	2 4445555	2 5555
2	2 66677777	2 667777
2 9	2 8888889999999	2 8888888899
3 0000001111	3 00000000111	3 0000000000001111
3 2222233	3 2223	3 22222333
3 4444444444555555	3 4	3 44
3 666777		3 6
3 888899		3 88
4 01		
4 2		
4 4		

The MINITAB output for the analysis of the iris data is given below.

**One-way ANOVA:**

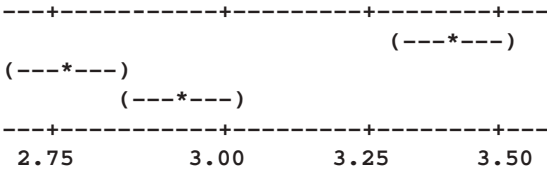
SOURCE	DF	SS	MS	F	P
IRIS	2	11.345	5.672	49.16	0.000
ERROR	147	16.962	0.115		
TOTAL	149	28.307			

S = 0.3397 R-Sq = 40.08%

LEVEL	N	MEAN	STDEV
1	50	3.4280	0.3791
2	50	2.7700	0.3138
3	50	2.9740	0.3225

POOLED STDEV = 0.3397

INDIVIDUAL 95% CIs FOR MEAN  
BASED ON POOLED STDEV



- (a) Identify the SSE and its degrees of freedom. Also locate  $s$ .
  - (b) Check the calculation of  $F$  from the given sums of squares and d.f.
  - (c) Is there one population with highest mean or are two or more alike? Use multiple- $t$  confidence intervals with  $\alpha = .05$ .
- 14.37 Three different chemicals are compared on their ability to make fabric stain-resistant. Four bolts of cloth, manufactured weeks apart, are used. Because the results are expected to vary from bolt of cloth to bolt, three small samples are cut from each bolt and a different chemical is applied to each. The assignment of chemicals is random. This process is repeated for each of the

four bolts. Suppose the measurement of stain-resistance are

	Block			
Treatment	1	2	3	4
1	8	9	1	6
2	5	12	0	11
3	8	15	8	13

- (a) Provide a decomposition for the observations from this randomized block experiment.
  - (b) Find the sum of squares for each array.
  - (c) Determine the degrees of freedom by checking the constraints for each array.
- 14.38 Refer to the output below concerning the time (min) it took four different persons (blocks) to complete three different tasks.
- (a) Identify the SSE and its degrees of freedom.
  - (b) Are the block means different? Check the calculation of  $F$  for blocks from the given sums of squares and degrees of freedom.
  - (c) Are the mean task times different? Check the calculation of  $F$  for treatments (tasks) from the given sums of squares and degrees of freedom.
  - (d) Use multiple- $t$  95% confidence intervals to investigate differences between mean task times.

**Two-way Analysis of Variance**

Analysis of Variance for Time					
Source	DF	SS	MS	F	P
Block	3	90.92	30.31	7.96	0.016
Treatment	2	71.17	35.58	9.35	0.014
Error	6	22.83	3.81		
Total	11	184.92			

Block	Mean	Treatment	Mean
1	9.0	1	8.00
2	13.7	2	9.50
3	6.7	3	13.75
4	12.3		