finish within the 2-hour period. To see whether this is actually the case, 120 students were randomly selected, and their completion times recorded. It was decided that $k = 8$ intervals should be used. The criteria imply that the 90th percentile of the completion time distribution is $\mu + 1.28\sigma = 120$. Since $\mu = 100$, this implies that $\sigma = 15.63$.

The eight intervals that divide the standard normal scale into eight equally likely segments are [0, .32), [.32, .675), [.675, 1.15), and [1.15, ∞), and their four counterparts are on the other side of 0. For $\mu = 100$ and $\sigma = 15.63$, these intervals become [100, 105), [105, 110.55), [110.55, 117.97), and [117.97, ∞). Thus $p_{i0} = \frac{1}{8} = .125$ $(i = 1, \ldots, 8)$, so each expected cell count is $np_{i0} = 120(.125) = 15$. The observed cell counts were 21, 17, 12, 16, 10, 15, 19, and 10, resulting in a χ^2 of 7.73. Since $\chi^2_{.10,7} = 12.017$ and 7.73 is not ≥ 12.017, there is no evidence for concluding that the criteria have not been met. ■

EXERCISES Section 14.1 (1–11)

1. What conclusion would be appropriate for an upper-tailed chi-squared test in each of the following situations?

a. $\alpha = .05$, df $= 4$, $\chi^2 = 12.25$
b. $\alpha = .01$, df $= 3$, $\chi^2 = 8.54$
c. $\alpha = .10$, df $= 2$, $\chi^2 = 4.36$
d. $\alpha = .01$, $k = 6$, $\chi^2 = 10.20$

2. Say as much as you can about the *P*-value for an upper-tailed chi-squared test in each of the following situations:

a. $\chi^2 = 7.5$, df $= 2$ **b.** $\chi^2 = 13.0$, df $= 6$
c. $\chi^2 = 18.0$, df $= 9$ **d.** $\chi^2 = 21.3$, df $= 5$
e. $\chi^2 = 5.0$, $k = 4$

3. The article "Racial Stereotypes in Children's Television Commercials" (*J. of Adver. Res.,* 2008: 80–93) reported the following frequencies with which ethnic characters appeared in recorded commercials that aired on Philadelphia television stations.

Ethnicity:	African American	Asian	Caucasian	Hispanic
Frequency:	57	11	330	6

The 2000 census proportions for these four ethnic groups are .177, .032, .734, and .057, respectively. Does the data suggest that the proportions in commercials are different from the census proportions? Carry out a test of appropriate hypotheses using a significance level of .01, and also say as much as you can about the *P*-value.

4. It is hypothesized that when homing pigeons are disoriented in a certain manner, they will exhibit no preference for any direction of flight after takeoff (so that the direction X should be uniformly distributed on the interval from 0° to 360°). To test this, 120 pigeons are disoriented, let loose, and the direction of flight of each is recorded; the resulting data follows. Use the chi-squared test at level .10 to see whether the data supports the hypothesis.

Direction	0−<45°	45−<90°	90−<135°
Frequency	12	16	17
Direction	135−< 180°	180−<225°	225−<270°
Frequency	15	13	20
Direction	270−<315°	315−<360°	
Frequency	17	10	

5. An information-retrieval system has ten storage locations. Information has been stored with the expectation that the long-run proportion of requests for location i is given by $p_i = (5.5 - |i - 5.5|)/30$. A sample of 200 retrieval requests gave the following frequencies for locations 1–10, respectively: 4, 15, 23, 25, 38, 31, 32, 14, 10, and 8. Use a chi-squared test at significance level .10 to decide whether the data is consistent with the *a priori* proportions (use the *P*-value approach).

6. The article "The Gap Between Wine Expert Ratings and Consumer Preferences" (*Intl. J. of Wine Business Res.,* 2008: 335–351) studied differences between expert and consumer ratings by considering medal ratings for wines, which could be gold (G), silver (S), or bronze (B). Three categories were then established: 1. Rating is the same [(G,G), (B,B), (S,S)]; 2. Rating differs by one medal [(G,S), (S,G), (S,B), (B,S)]; and 3. Rating differs by two medals [(G,B), (B,G)]. The observed frequencies for these three categories were 69, 102, and 45, respectively. On the hypothesis of equally likely expert ratings and consumer ratings being assigned completely by chance, each of the nine medal pairs has probability 1/9. Carry out an appropriate chi-squared test using a significance level of .10 by first obtaining *P*-value information.

7. Criminologists have long debated whether there is a relationship between weather conditions and the incidence of violent crime. The author of the article "Is There a Season for Homicide?" (*Criminology,* 1988: 287–296) classified 1361 homicides according to season, resulting in the accompanying data. Test the null hypothesis of equal proportions using $\alpha = .01$ by using the chi-squared table to say as much as possible about the *P*-value.

Winter	Spring	Summer	Fall
328	334	372	327

8. The article "Psychiatric and Alcoholic Admissions Do Not Occur Disproportionately Close to Patients' Birthdays" (*Psychological Reports,* 1992: 944–946) focuses on the existence of any relationship between the date of patient admission for treatment of alcoholism and the patient's birthday. Assuming a 365-day year (i.e., excluding leap year), in the absence of any relation, a patient's admission date is equally likely to be any one of the 365 possible days. The investigators established four different admission categories: (1) within 7 days of birthday; (2) between 8 and 30 days, inclusive, from the birthday; (3) between 31 and 90 days, inclusive, from the birthday; and (4) more than 90 days from the birthday. A sample of 200 patients gave observed frequencies of 11, 24, 69, and 96 for categories 1, 2, 3, and 4, respectively. State and test the relevant hypotheses using a significance level of .01.

9. The response time of a computer system to a request for a certain type of information is hypothesized to have an exponential distribution with parameter $\lambda = 1$ sec (so if X = response time, the pdf of X under H_0 is $f_0(x) = e^{-x}$ for $x \geq 0$).

a. If you had observed $X_1, X_2, \ldots, X_n$ and wanted to use the chi-squared test with five class intervals having equal probability under H_0, what would be the resulting class intervals?

b. Carry out the chi-squared test using the following data resulting from a random sample of 40 response times:

.10	.99	1.14	1.26	3.24	.12	.26	.80
.79	1.16	1.76	.41	.59	.27	2.22	.66
.71	2.21	.68	.43	.11	.46	.69	.38
.91	.55	.81	2.51	2.77	.16	1.11	.02
2.13	.19	1.21	1.13	2.93	2.14	.34	.44

10. a. Show that another expression for the chi-squared statistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{N_i^2}{np_{i0}} - n$$

Why is it more efficient to compute χ^2 using this formula?

b. When the null hypothesis is (H_0: $p_1 = p_2 = \cdots = p_k = 1/k$ (i.e., $p_{i0} = 1/k$ for all i), how does the formula of part (a) simplify? Use the simplified expression to calculate χ^2 for the pigeon/direction data in Exercise 4.

11. a. Having obtained a random sample from a population, you wish to use a chi-squared test to decide whether the population distribution is standard normal. If you base the test on six class intervals having equal probability under H_0, what should be the class intervals?

b. If you wish to use a chi-squared test to test H_0: the population distribution is normal with $\mu = .5$, $\sigma = .002$ and the test is to be based on six equiprobable (under H_0) class intervals, what should be these intervals?

c. Use the chi-squared test with the intervals of part (b) to decide, based on the following 45 bolt diameters, whether bolt diameter is a normally distributed variable with $\mu = .5$ in., $\sigma = .002$ in.

.4974	.4976	.4991	.5014	.5008	.4993
.4994	.5010	.4997	.4993	.5013	.5000
.5017	.4984	.4967	.5028	.4975	.5013
.4972	.5047	.5069	.4977	.4961	.4987
.4990	.4974	.5008	.5000	.4967	.4977
.4992	.5007	.4975	.4998	.5000	.5008
.5021	.4959	.5015	.5012	.5056	.4991
.5006	.4987	.4968			

14.2 Goodness-of-Fit Tests for Composite Hypotheses

In the previous section, we presented a goodness-of-fit test based on a χ^2 statistic for deciding between H_0: $p_1 = p_{10}, \ldots, p_k = p_{k0}$ and the alternative H_a stating that H_0 is not true. The null hypothesis was a **simple hypothesis** in the sense that each p_{i0} was a specified number, so that the expected cell counts when H_0 was true were uniquely determined numbers.

In many situations, there are k naturally occurring categories, but H_0 states only that the p_i's are functions of other parameters $\theta_1, \ldots, \theta_m$ without specifying the values of these θ's. For example, a population may be in equilibrium with respect to proportions of the three genotypes *AA*, *Aa*, and *aa*. With p_1, p_2, and p_3 denoting these proportions (probabilities), one may wish to test

$$H_0: p_1 = \theta^2,\ p_2 = 2\theta(1 - \theta),\ p_3 = (1 - \theta)^2 \qquad (14.1)$$

Monographs, 1971: 129–152) presents the accompanying data on the variable $X =$ the number of hops before the first flight and preceded by a flight. The author then proposed and fit a geometric probability distribution $[p(x) = P(X = x) = p^{x-1} \cdot q$ for $x = 1, 2, \ldots,$ where $q = 1 - p]$ to the data. The total sample size was $n = 130$.

x	1	2	3	4	5	6	7	8	9	10	11	12
Number of Times x Observed	48	31	20	9	6	5	4	2	1	1	2	1

a. The likelihood is $(p^{x_1-1} \cdot q) \cdot \cdots \cdot (p^{x_n-1} \cdot q) = p^{\Sigma x_i - n} \cdot q^n$. Show that the mle of p is given by $\hat{p} = (\Sigma x_i - n)/\Sigma x_i$, and compute $\hat{p}$ for the given data.

b. Estimate the expected cell counts using $\hat{p}$ of part (a) [expected cell counts $= n \cdot (\hat{p})^{x-1} \cdot \hat{q}$ for $x = 1, 2, \ldots$], and test the fit of the model using a χ^2 test by combining the counts for $x = 7, 8, \ldots,$ and 12 into one cell ($x \geq 7$).

15. A certain type of flashlight is sold with the four batteries included. A random sample of 150 flashlights is obtained, and the number of defective batteries in each is determined, resulting in the following data:

Number Defective	0	1	2	3	4
Frequency	26	51	47	16	10

Let X be the number of defective batteries in a randomly selected flashlight. Test the null hypothesis that the distribution of X is Bin(4, θ). That is, with $p_i = P(i$ defectives), test

$$H_0: p_i = \binom{4}{i} \theta^i (1 - \theta)^{4-i} \qquad i = 0, 1, 2, 3, 4$$

[*Hint:* To obtain the mle of θ, write the likelihood (the function to be maximized) as $\theta^u(1 - \theta)^v$, where the exponents u and v are linear functions of the cell counts. Then take the natural log, differentiate with respect to θ, equate the result to 0, and solve for $\hat{\theta}$.]

16. In a genetics experiment, investigators looked at 300 chromosomes of a particular type and counted the number of sister-chromatid exchanges on each ("On the Nature of Sister-Chromatid Exchanges in 5-Bromodeoxyuridine-Substituted Chromosomes," *Genetics,* 1979: 1251–1264). A Poisson model was hypothesized for the distribution of the number of exchanges. Test the fit of a Poisson distribution to the data by first estimating μ and then combining the counts for $x = 8$ and $x = 9$ into one cell.

x = Number of Exchanges	0	1	2	3	4	5	6	7	8	9
Observed Counts	6	24	42	59	62	44	41	14	6	2

17. An article in *Annals of Mathematical Statistics* reports the following data on the number of borers in each of 120 groups of borers. Does the Poisson pmf provide a plausible model for the distribution of the number of borers in a group? [*Hint:* Add the frequencies for 7, 8, . . . , 12 to establish a single category "≥ 7."]

Number of Borers	0	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	24	16	16	18	15	9	6	5	3	4	3	0	1

18. The article "A Probabilistic Analysis of Dissolved Oxygen–Biochemical Oxygen Demand Relationship in Streams" (*J. Water Resources Control Fed.,* 1969: 73–90) reports data on the rate of oxygenation in streams at 20°C in a certain region. The sample mean and standard deviation were computed as $\bar{x} = .173$ and $s = .066$, respectively. Based on the accompanying frequency distribution, can it be concluded that oxygenation rate is a normally distributed variable? Use the chi-squared test with $\alpha = .05$.

Rate (per day)	**Frequency**
Below .100	12
.100–below .150	20
.150–below .200	23
.200–below .250	15
.250 or more	13

19. Each headlight on an automobile undergoing an annual vehicle inspection can be focused either too high (H), too low (L), or properly (N). Checking the two headlights simultaneously (and not distinguishing between left and right) results in the six possible outcomes HH, LL, NN, HL, HN, and LN. If the probabilities (population proportions) for the single headlight focus direction are $P(H) = \theta_1$, $P(L) = \theta_2$, and $P(N) = 1 - \theta_1 - \theta_2$ and the two headlights are focused independently of one another, the probabilities of the six outcomes for a randomly selected car are the following:

$$p_1 = \theta_1^2 \quad p_2 = \theta_2^2 \quad p_3 = (1 - \theta_1 - \theta_2)^2$$

$$p_4 = 2\theta_1\theta_2 \quad p_5 = 2\theta_1(1 - \theta_1 - \theta_2)$$

$$p_6 = 2\theta_2(1 - \theta_1 - \theta_2)$$

Use the accompanying data to test the null hypothesis

$$H_0: p_1 = \pi_1(\theta_1, \theta_2), \ldots, p_6 = \pi_6(\theta_1, \theta_2)$$

where the $\pi_i(\theta_1, \theta_2)$s are given previously.

Outcome	*HH*	*LL*	*NN*	*HL*	*HN*	*LN*
Frequency	49	26	14	20	53	38

[*Hint:* Write the likelihood as a function of θ_1 and θ_2, take the natural log, then compute $\partial/\partial\theta_1$ and $\partial/\partial\theta_2$, equate them to 0, and solve for $\hat{\theta}_1, \hat{\theta}_2$.]

20. The article "Compatibility of Outer and Fusible Interlining Fabrics in Tailored Garments (*Textile Res. J.,* 1997: 137–142) gave the following observations on bending

Models and methods for analyzing data in which each individual is categorized with respect to three or more factors (multidimensional contingency tables) are discussed in several of the chapter references.

EXERCISES Section 14.3 (24–36)

24. The accompanying two-way table was constructed using data in the article "Television Viewing and Physical Fitness in Adults" (*Research Quarterly for Exercise and Sport,* 1990: 315–320). The author hoped to determine whether time spent watching television is associated with cardiovascular fitness. Subjects were asked about their television-viewing habits and were classified as physically fit if they scored in the excellent or very good category on a step test. We include Minitab output from a chi-squared analysis. The four TV groups corresponded to different amounts of time per day spent watching TV (0, 1–2, 3–4, or 5 or more hours). The 168 individuals represented in the first column were those judged physically fit. Expected counts appear below observed counts, and Minitab displays the contribution to χ^2 from each cell. State and test the appropriate hypotheses using $\alpha = .05$.

```
               1           2       Total
1             35         147         182
           25.48      156.52
2            101         629         730
          102.20      627.80
3             28         222         250
           35.00      215.00
4              4          34          38
            5.32       32.68
Total        168        1032        1200

ChiSq = 3.557 + 0.579 +
        0.014 + 0.002 +
        1.400 + 0.228 +
        0.328 + 0.053 = 6.161

df = 3
```

25. The accompanying data refers to leaf marks found on white clover samples selected from both long-grass areas and short-grass areas ("The Biology of the Leaf Mark Polymorphism in *Trifolium repens* L.," *Heredity,* 1976: 306–325). Use a χ^2 test to decide whether the true proportions of different marks are identical for the two types of regions.

	Type of Mark					Sample Size
	L	LL	Y + YL	O	Others	
Long-Grass Areas	409	11	22	7	277	726
Short-Grass Areas	512	4	14	11	220	761

26. The following data resulted from an experiment to study the effects of leaf removal on the ability of fruit of a certain type to mature ("Fruit Set, Herbivory, Fruit Reproduction, and the Fruiting Strategy of *Catalpa speciosa,*" *Ecology,* 1980: 57–64):

Treatment	Number of Fruits Matured	Number of Fruits Aborted
Control	141	206
Two leaves removed	28	69
Four leaves removed	25	73
Six leaves removed	24	78
Eight leaves removed	20	82

Does the data suggest that the chance of a fruit maturing is affected by the number of leaves removed? State and test the appropriate hypotheses at level .01.

27. The article "Human Lateralization from Head to Foot: Sex-Related Factors" (*Science,* 1978: 1291–1292) reports for both a sample of right-handed men and a sample of right-handed women the number of individuals whose feet were the same size, had a bigger left than right foot (a difference of half a shoe size or more), or had a bigger right than left foot.

	L > R	L = R	L < R	Sample Size
Men	2	10	28	40
Women	55	18	14	87

Does the data indicate that gender has a strong effect on the development of foot asymmetry? State the appropriate null and alternative hypotheses, compute the value of χ^2, and obtain information about the *P*-value.

28. A random sample of 175 Cal Poly State University students was selected, and both the email service provider and cell phone provider were determined for each one, resulting in the accompanying data. State and test the appropriate hypotheses using the *P*-value approach.

		Cell Phone Provider		
		ATT	Verizon	Other
Email Provider	**gmail**	28	17	7
	Yahoo	31	26	10
	Other	26	19	11

29. The accompanying data on degree of spirituality for samples of natural and social scientists at research universities as well as for a sample of non-academics with graduate degrees appeared in the article "Conflict Between Religion and Science Among Academic Scientists" (*J. for the Scientific Study of Religion,* 2009: 276–292).

	Degree of Spirituality			
	Very	**Moderate**	**Slightly**	**Not at all**
N.S.	56	162	198	211
S.S.	56	223	243	239
G.D.	109	164	74	28

a. Is there substantial evidence for concluding that the three types of individuals are not homogenous with respect to their degree of spirituality? State and test the appropriate hypotheses.

b. Considering just the natural scientists and social scientists, is there evidence for non-homogeneity? Base your conclusion on a P-value.

30. Three different design configurations are being considered for a particular component. There are four possible failure modes for the component. An engineer obtained the following data on number of failures in each mode for each of the three configurations. Does the configuration appear to have an effect on type of failure?

		Failure Mode			
		1	**2**	**3**	**4**
Configuration	**1**	20	44	17	9
	2	4	17	7	12
	3	10	31	14	5

31. A random sample of smokers was obtained, and each individual was classified both with respect to gender and with respect to the age at which he/she first started smoking. The data in the accompanying table is consistent with summary results reported in the article "Cigarette Tar Yields in Relation to Mortality in the Cancer Prevention Study II Prospective Cohort" (*British Med. J.,* 2004: 72–79).

		Gender	
		Male	**Female**
Age	**<16**	25	10
	16 − 17	24	32
	18 − 20	28	17
	>20	19	34

a. Calculate the proportion of males in each age category, and then do the same for females. Based on these proportions, does it appear that there might be an association between gender and the age at which an individual first smokes?

b. Carry out a test of hypotheses to decide whether there is an association between the two factors.

32. Each individual in a random sample of high school and college students was cross-classified with respect to both political views and marijuana usage, resulting in the data displayed in the accompanying two-way table ("Attitudes About Marijuana and Political Views," *Psychological Reports,* 1973: 1051–1054). Does the data support the hypothesis that political views and marijuana usage level are independent within the population? Test the appropriate hypotheses using level of significance .01.

		Usage Level		
		Never	**Rarely**	**Frequently**
Political Views	**Liberal**	479	173	119
	Conservative	214	47	15
	Other	172	45	85

33. Show that the chi-squared statistic for the test of independence can be written in the form

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{N_{ij}^2}{\hat{E}_{ij}}\right) - n$$

Why is this formula more efficient computationally than the defining formula for χ^2?

34. Suppose that in Exercise 32 each student had been categorized with respect to political views, marijuana usage, and religious preference, with the categories of this latter factor being Protestant, Catholic, and other. The data could be displayed in three different two-way tables, one corresponding to each category of the third factor. With $p_{ijk} = P$(political category i, marijuana category j, and religious category k), the null hypothesis of independence of all three factors states that $p_{ijk} = p_{i..}p_{.j.}p_{..k}$. Let n_{ijk} denote the observed frequency in cell (i, j, k). Show how to estimate the expected cell counts assuming that H_0 is true ($\hat{e}_{ijk} = n\hat{p}_{ijk}$, so the $\hat{p}_{ijk}$'s must be determined). Then use the general rule of thumb to determine the number of degrees of freedom for the chi-squared statistic.

35. Suppose that in a particular state consisting of four distinct regions, a random sample of n_k voters is obtained from the kth region for $k = 1, 2, 3, 4$. Each voter is then classified according to which candidate (1, 2, or 3) he or she prefers and according to voter registration (1 = Dem., 2 = Rep., 3 = Indep.). Let p_{ijk} denote the proportion of voters in region k who belong in candidate category i and registration category j. The null hypothesis of homogeneous regions is H_0: $p_{ij1} = p_{ij2} = p_{ij3} = p_{ij4}$ for all i, j (i.e., the proportion within each candidate/registration combination is the same for all four regions). Assuming that H_0 is true, determine $\hat{p}_{ijk}$ and $\hat{e}_{ijk}$ as functions of the observed n_{ijk}'s, and use the general rule of thumb to obtain the number of degrees of freedom for the chi-squared test.

36. Consider the accompanying 2 × 3 table displaying the sample proportions that fell in the various combinations of categories (e.g., 13% of those in the sample were in the first category of both factors).

	1	2	3
1	.13	.19	.28
2	.07	.11	.22

a. Suppose the sample consisted of $n = 100$ people. Use the chi-squared test for independence with significance level .10.

b. Repeat part (a), assuming that the sample size was $n = 1000$.

c. What is the smallest sample size n for which these observed proportions would result in rejection of the independence hypothesis?

SUPPLEMENTARY EXERCISES (37–49)

37. The article "Birth Order and Political Success" (*Psych. Reports,* 1971: 1239–1242) reports that among 31 randomly selected candidates for political office who came from families with four children, 12 were firstborn, 11 were middle born, and 8 were last born. Use this data to test the null hypothesis that a political candidate from such a family is equally likely to be in any one of the four ordinal positions.

38. Does the phase of the moon have any bearing on birthrate? Each of 222,784 births that occurred during a period encompassing 24 full lunar cycles was classified according to lunar phase. The following data is consistent with summary quantities that appeared in the article "The Effect of the Lunar Cycle on Frequency of Births and Birth Complications" (*Amer. J. of Obstetrics and Gynecology,* 2005: 1462–1464).

Lunar Phase	# Days in Phase	# Births
New moon	24	7680
Waxing crescent	152	48,442
First quarter	24	7579
Waxing gibbous	149	47,814
Full moon	24	7711
Waning gibbous	150	47,595
Last quarter	24	7733
Waning crescent	152	48,230

State and test the appropriate hypotheses to answer the question posed at the beginning of this exercise.

39. Qualifications of male and female head and assistant college athletic coaches were compared in the article "Sex Bias and the Validity of Believed Differences Between Male and Female Interscholastic Athletic Coaches" (*Research Quarterly for Exercise and Sport,* 1990: 259–267). Each person in random samples of 2225 male coaches and 1141 female coaches was classified according to number of years of coaching experience to obtain the accompanying two-way table. Is there enough evidence to conclude that the proportions falling into the experience categories are different for men and women? Use $\alpha = .01$.

	Years of Experience				
Gender	**1–3**	**4–6**	**7–9**	**10–12**	**13+**
Male	202	369	482	361	811
Female	230	251	238	164	258

40. The authors of the article "Predicting Professional Sports Game Outcomes from Intermediate Game Scores" (*Chance,* 1992: 18–22) used a chi-squared test to determine whether there was any merit to the idea that basketball games are not settled until the last quarter, whereas baseball games are over by the seventh inning. They also considered football and hockey. Data was collected for 189 basketball games, 92 baseball games, 80 hockey games, and 93 football games. The games analyzed were sampled randomly from all games played during the 1990 season for baseball and football and for the 1990–1991 season for basketball and hockey. For each game, the late-game leader was determined, and then it was noted whether the late-game leader actually ended up winning the game. The resulting data is summarized in the accompanying table.

Sport	Late-Game Leader Wins	Late-Game Leader Loses
Basketball	150	39
Baseball	86	6
Hockey	65	15
Football	72	21