CHAPTER 13

Chi-Squared Goodness-of-Fit Tests

Still, it is an error to argue in front of your data. You find yourself insensibly twisting them round to fit your theories.

Sherlock Holmes, The Adventures of Wisteria Lodge

A few observations and much reasoning lead to error; many observations and a little reasoning to truth.

Alexis Carrel

CONTENTS

We consider a population in which each member has any one of k possible values. We show how to test the hypothesis that a specified set of probabilities represents the proportions of the members of the population that have each of the different possible values. We consider populations in which each member is classified as

Introductory Statistics, DOI: 10.1016/B978-0-12-374388-6.00013-2

having two values, and we show how to test the hypothesis that the two values of a randomly selected member of the population are independent.

13.1 INTRODUCTION

The manipulation of data to make them conform to a particular scientific hypothesis is considered to be an instance of scientific fraud. There have been many such cases of scientific fraud over the years, ranging in severity from slight "fudging" to outright falsification of data. For instance, one of the most egregious examples involved the British educational psychologist Cyril Burt. Burt was highly regarded in his lifetime—indeed, he was eventually knighted by the Queen of England and became Sir Cyril—for his research on the IQs of identical twins who were raised apart. However, it is now widely accepted that in his published work he invented not only the data he published but also the very existence of his supposed research subjects and collaborators.

Perhaps the most puzzling instance of scientific fraud involves the Austrian monk Gregor Mendel (1822–1884), who is regarded as the founder of the theory of genetics. In 1865 Mendel published a paper outlining the results of a series of experiments carried out on garden peas. One of the experiments was concerned with the color—either yellow or green—of the seeds of such peas. Mendel began his experiment by breeding peas of pure yellow strain, which is a strain of peas in which every plant in every generation has only yellow seeds. He also bred a pure green strain. Mendel then crossed peas of the pure yellow strain with those of the pure green strain. The result of this crossing, known as *first-generation hybrid* seeds, was always a yellow seed. That is, there were no green seeds in this generation.

Mendel then crossed these first-generation seeds with themselves, to obtain second-generation seeds. Surprisingly, green seeds reappeared in this generation. In fact, approximately 25 percent of second-generation seeds were green, and 75 percent were yellow.

In his paper, Mendel presented a theory to explain these results. His theory supposed that each seed contained two entities, which we now call *genes*, that together determine the color of the seed. Each gene is one of two types: type y (for yellow) or type g (for green). Mendel's theory was that the pair of genes in the pure yellow strain seeds is always y,y. That is, both genes in a pea from the pure yellow strain are yellow. Similarly, the pair of genes in seeds from the pure green strain are g,g. Mendel now supposed that when two seeds are crossed, the resulting offspring obtains one gene from each parent. In addition, Mendel supposed that the gene obtained from a parent is equally likely to be either one of the two genes of that parent. Thus, when a pure y,y yellow seed is matched with a pure g,g green seed, the offspring will necessarily have one y and one g seed; that is, the offspring will have the gene pair y,g. Since every offspring resulting from a cross of a pure yellow and a pure green seed was itself yellow, Mendel postulated that the y gene was dominant over the g gene, in that a seed having the gene pair y,g would be yellow. See Fig. 13.1.

Consider now what happens when two first-generation seeds are crossed. First note that both seeds are hybrids having the gene pair y,g. Also note that in order for the offspring seed to be green, it must receive the g gene from each parent. Since each parent is equally likely to contribute either its y or its g gene, it follows that the probability that both parents contribute their g genes is $1/2 \times 1/2 = 1/4$. Thus, the result of a large number of crossings of first-generation seeds should be that approximately 25 percent of the next-generation seeds are green. This is exactly the result reported by Mendel. See Fig. 13.2.

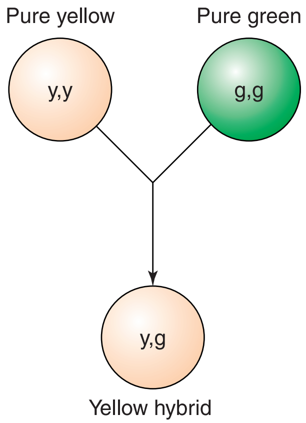

FIGURE 13.1

Crossing pure yellow seeds with pure green seeds.

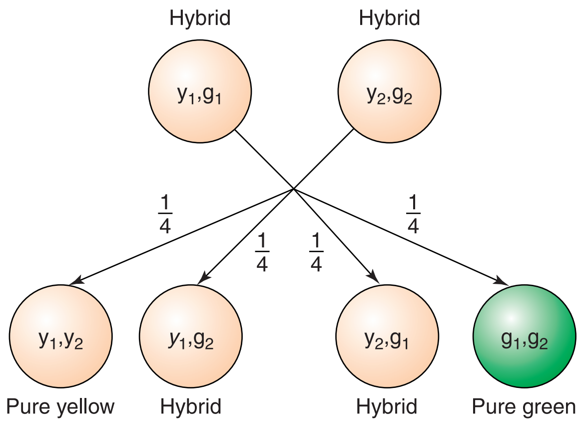

FIGURE 13.2

Crossing hybrid first-generation seeds.

Mendel's discovery is first-rate science of the highest order. Although it took some time (as his work was ignored for almost 30 years), his theory of genetics has become a cornerstone of basic science.

It is not difficult to imagine the consternation among geneticists when, in 1936, R. A. Fisher published a paper that analyzed Mendel's data and concluded that they fit the theory too well to be explained by chance. Using the chi-squared goodness-of-fit test that had been developed by Karl Pearson, Fisher showed that an overall data fit at least as good as the one reported by Mendel would have occurred with a probability equal to 0.00004.

For instance, Mendel reported that of 8023 second-generation peas, 6022 were yellow and 2001 were green. That is, the fraction of the second-generation peas that was green was $2001/8023 = 0.2494$, which is almost exactly equal to the theoretical probability 0.25. Although such a good fit in itself is not that unlikely (at least as good a fit as this would occur roughly 10 percent of the time), the trouble was that almost all the experiments reported by Mendel resulted in data that were in unusually close agreement with the theoretical probability. By combining the results of all the experiments reported by Mendel, Fisher came up with the (p value) probability of 0.00004.

Although believing that Mendel's data had been manipulated, Fisher apparently exonerated Mendel himself from direct blame. Indeed, Fisher has gone on record as believing that the data were probably manipulated by an assistant who knew the results that Mendel expected. (Of course, equally plausible is that Mendel himself made mistakes in recording the data. Even honest people can see what is not there, when they believe it should be.)

In Sec. 13.2 we present the chi-squared goodness-of-fit test, which can be used to determine how well a given data set fits a particular probability model. Its use enables us to test the validity of the probability model.

In Sec. 13.3 we consider populations in which each member is classified according to two distinct characteristics. We show how the goodness-of-fit test can be used to test the hypothesis that the two characteristics of a randomly chosen member of the population are independent.

The two characteristics of members of a population will be independent if knowledge of one of the characteristics of a randomly chosen member of the population does not affect the probabilities of the other characteristic of this member. Whereas in Sec. 13.3 we suppose that the data result from a random sample of the entire population, in Sec. 13.4 we consider a different type of sampling scheme. This new scheme starts by focusing attention on one of the characteristics. It determines the various possible values of this characteristic and then chooses random samples from the subpopulations of members having each of the possible values. For instance, if one of the characteristics is gender, then rather than choose a random sample from the entire population, as is done in Sec. 13.3, we now choose random samples from the subpopulations of men and women. A test for independence is presented when this type of sampling scheme is used. In addition, we show how the results of Sec. 13.4 can be used to test the hypothesis that an arbitrary number of population proportions are equal. In the special case of two populations, the test is identical to the one presented in Sec. 10.6.

13.2 CHI-SQUARED GOODNESS-OF-FIT TESTS

Consider a very large population, and suppose that each member of the population has a value that can be 1 or 2 or 3 or ... or k. For a given set of probabilities $p_i, i = 1, \ldots, k$, we will consider the problem of testing the null hypothesis that p_i represents, for each i, the proportion of the population that has value i. That is, if we let P_i denote the true proportion of the population that has value i, for $i = 1, \ldots, k$, then we are interested in testing

$$\mathrm{H}_0\colon P_1 = p_1,\ P_2 = p_2, \ldots, P_k = p_k$$

against the alternative hypothesis

$$\mathrm{H}_1\colon P_i \neq p_i \qquad \text{for some } i,\ i = 1, \ldots, k$$

■ Example 13.1

It is known that 41 percent of the U.S. population has type A blood, 9 percent has type B, 4 percent has type AB, and 46 percent has type O. Suppose that we suspect that the blood type distribution of people suffering from stomach cancer is different from that of the overall population.

To verify that the blood type distribution is different for those suffering from stomach cancer, we could test the null hypothesis

$$\mathrm{H}_0\colon P_1 = 0.41,\ P_2 = 0.09,\ P_3 = 0.04,\ P_4 = 0.46$$

where P_1 is the proportion of all those with stomach cancer who have type A blood, P_2 is the proportion of those who have type B blood, P_3 is the proportion who have type AB blood, and P_4 is the proportion who have type O blood. A rejection of H_0 would then enable us to conclude that the blood type distribution is indeed different for those suffering from stomach cancer.

In the preceding scenario, each member of the population of individuals who are suffering from stomach cancer is given one of four possible values according to his or her blood type. We are interested in testing the hypothesis that $P_1 = 0.41, P_2 = 0.09, P_3 = 0.04, P_4 = 0.46$ represent the proportions of this population having each of the different values. ■

To test the null hypothesis that $P_i = p_i, i = 1, \ldots, k$, first we need to draw a random sample of elements from the population. Suppose this sample is of size n. Let N_i denote the number of elements of the sample that have value i, for $i = 1, \ldots, k$. Now, if the null hypothesis is true, then each element of the sample will have value i with probability p_i. Also, since the population is assumed to be very large, it follows that the successive values of the members of the sample will be independent. Thus, if the null hypothesis is true, then N_i will have the same distribution as the number of successes in n independent trials, when each trial is a success with probability p_i. That is, if H_0 is true, then N_i will be a binomial random variable with parameters n and p_i. Since the expected value of a binomial is the product of its parameters, we see that when H_0 is true,

$$E[N_i] = np_i \qquad i = 1, \ldots, k$$

For each i, let e_i denote this expected number of outcomes that equal i when H_0 is true. That is,

$$e_i = np_i$$

Thus, when H_0 is true, we expect that N_i would be relatively close to e_i. That is, when the null hypothesis is true, the quantity $(N_i - e_i)^2$ should not be too large, say, in relation to e_i. Since this is true for each value of i, a reasonable way of testing H_0 would be to compute the value of the test statistic

$$\mathrm{TS} = \sum_{i=1}^{k} \frac{(N_i - e_i)^2}{e_i}$$

and then reject H_0 when TS is sufficiently large.

To determine how large TS need be to justify rejection of the null hypothesis, we use a result that was proved by Karl Pearson in 1900. This result states that for large values of n, TS will have an approximately chi-squared distribution with $k - 1$ degrees of freedom. Let $\chi^2_{k-1,\alpha}$ denote the $100(1 - \alpha)$th percentile of this distribution; that is, a chi-squared random variable having $k - 1$ degrees of freedom will exceed this value with probability α (Fig. 13.3). Then the approximate significance-level-α test of the null hypothesis H_0 against the alternative H_1 is as follows:

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } \mathrm{TS} \geq \chi^2_{k-1,\alpha} \\ \text{Do not reject } H_0 & \text{otherwise} \end{array}$$

The preceding is called the *chi-squared goodness-of-fit test*. For reasonably large values of n, it results in a hypothesis test of H_0 whose significance level is approximately equal to α. An accepted rule of thumb is that this approximation will be quite good provided n is large enough so that $e_i \geq 1$ for each i and at least 80 percent of the values e_i exceed 5.

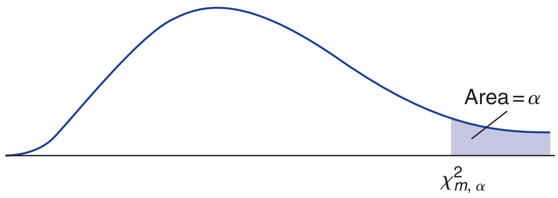

FIGURE 13.3
Chi-squared percentile $P(\chi^2_m \geq \chi^2_{m,\alpha}) = \alpha$.

Table 13.1 Some Values of $\chi^2_{m,\alpha}$

m	$\alpha = 0.99$	$\alpha = 0.95$	$\alpha = 0.05$	$\alpha = 0.01$
1	0.000157	0.00393	3.841	6.635
2	0.0201	0.103	5.991	9.210
3	0.115	0.352	7.815	11.345
4	0.297	0.711	9.488	13.277
5	0.554	1.145	11.070	15.086
6	0.872	1.635	12.592	16.812
7	1.239	2.167	14.067	18.475

Values of $\chi^2_{m,\alpha}$ for various values of m and α are given in App. Table D.3. A portion of this table is represented in Table 13.1.

Example 13.2

Suppose, in Example 13.1, that a random sample of 200 stomach cancer patients yielded 92 having blood type A, 20 having blood type B, 4 having blood type AB, and 84 having blood type O. Are these data significant enough, at the 5 percent level of significance, to enable us to reject the null hypothesis that the blood type distribution of stomach cancer sufferers is the same as that of the general population?

Solution

The observed frequencies are

$$N_1 = 92 \quad N_2 = 20 \quad N_3 = 4 \quad N_4 = 84$$

whereas the expected frequencies, when H_0 is true, are

$$e_1 = np_1 = 200 \times 0.41 = 82$$
$$e_2 = np_2 = 200 \times 0.09 = 18$$
$$e_3 = np_3 = 200 \times 0.04 = 8$$
$$e_4 = np_4 = 200 \times 0.46 = 92$$

Thus, the value of the test statistic is

$$\begin{aligned}\text{TS} &= \frac{(92-82)^2}{82} + \frac{(20-18)^2}{18} + \frac{(4-8)^2}{8} + \frac{(84-92)^2}{92} \\ &= 4.1374\end{aligned}$$

Since this value is not as large as $\chi^2_{3,\,0.05} = 7.815$ (obtained from Table 13.1), it follows that we cannot reject, at the 5 percent level of significance, the null hypothesis that the blood type distribution of people with stomach cancer is the same as that of the general public. ■

The chi-squared goodness-of-fit test can also be performed by determining the p value of the resulting data. If the data result in the test statistic having a value v, then the p value equals the probability that a value at least as large as v will have occurred if H_0 is true. Now, when H_0 is true, the distribution of the test statistic TS is approximately chi squared with $k-1$ degrees of freedom. Thus, it follows that the p value is approximately equal to the probability that a chi-squared random variable with $k-1$ degrees of freedom is at least as large as v. The null hypothesis is then rejected at any significance level greater than or equal to the p value and is not rejected at all lower significance levels.

To Determine the p value of the Chi-Squared Test

1. Calculate the value of the test statistic TS.
2. If the value of TS is v, then the p value is

$$p \text{ value} = P\{\chi^2_{k-1} \geq v\}$$

where χ^2_{k-1} is a chi-squared random variable with $k-1$ degrees of freedom.

Program 13-1 can be used to determine both the value of the test statistic TS and the resulting p value.

■ Example 13.3

To determine whether accidents are more likely to occur on certain days of the week, data have been collected on all the accidents requiring medical attention that occurred over the last 12 months at an automobile plant in northern California. The data yielded a total of 250 accidents, with the number occurring on each day of the week being as follows:

Monday	62
Tuesday	47
Wednesday	44
Thursday	45
Friday	52

Use the preceding data to test, at the 5 percent level of significance, the hypothesis that an accident is equally likely to occur on any day of the week.

Solution

We want to test the null hypothesis that

$$P_i = \frac{1}{5} \qquad i = 1, 2, 3, 4, 5$$

The observed data are $N_1 = 62$, $N_2 = 47$, $N_3 = 44$, $N_4 = 45$, and $N_5 = 52$. Running Program 13-1 yields for the value of TS and the resulting p value

$$\text{TS} = 4.36 \qquad p \text{ value} = 0.359$$

Thus, a value of TS at least as large as the one obtained would be expected to occur 35.9 percent of the time when H_0 is true, and so the null hypothesis that accidents are equally likely to occur on any day of the week cannot be rejected. ■

Sometimes a data set is reported that is in such strong agreement with the expectations of the null hypothesis that one becomes suspicious about the possibility that the data may have been manipulated. One way of ascertaining the likelihood of this possibility is to calculate the value v of the test statistic TS and then to determine how likely it is that a value as small as or smaller than v will have occurred when the null hypothesis is true. That is, one should determine $P\{\chi^2_{k-1} \geq v\}$. An extremely small value of this probability is then strong evidence for possible data manipulation.

■ Example 13.4

In the introduction to this chapter, we commented on an experiment performed by Gregor Mendel in which he reported that a cross of 8023 hybrid peas resulted in 6022 yellow and 2001 green peas. In theory, each cross should result in a yellow pea with probability 3/4 and a green one with probability 1/4. To determine if the data fit the model too well, we start by determining the value of the test statistic TS.

The parameters of this problem are

$$n = 8023 \quad k = 2 \quad p_1 = 3/4 \quad p_2 = 1/4 \quad N_1 = 6022 \quad N_2 = 2001$$

Since

$$e_1 = 8023 \times \frac{3}{4} = 6017.25$$
$$e_2 = 8023 \times \frac{1}{4} = 2005.75$$

the value of the test statistic TS is

$$\begin{aligned}\text{TS} &= \frac{(6022 - 6017.25)^2}{6017.25} + \frac{(2001 - 2005.75)^2}{2005.75} \\ &= 0.015\end{aligned}$$

Since 0.015 is greater than $\chi^2_{1,0.95} = 0.004$, it follows that a value as small as or smaller than 0.015 would occur over 5 percent of the time (see Fig. 13.4). Thus, the data do not indicate any manipulation.

Indeed, it can be computed (say, from Program 13-2, which would give the p value $P\{\chi^2_1 \geq 0.015\}$ as output) that

$$P\left\{\chi^2_1 \leq 0.015\right\} = 0.0974$$

and so roughly 10 percent of the time TS would be as small as the value obtained when Mendel's data were used. While this by itself is not suggestive of any (conscious or unconscious) data manipulation, it turns out that almost all the data sets reported by Mendel fit his theoretical expectations as well as this one. Indeed, the probability that the sum of the values of all the chi-squared test statistics reported by Mendel, one for each experiment, would be as small as or smaller than the value obtained by using Mendel's data is 0.00004. ■

Table 13.2 summarizes the chi-squared goodness-of-fit test.

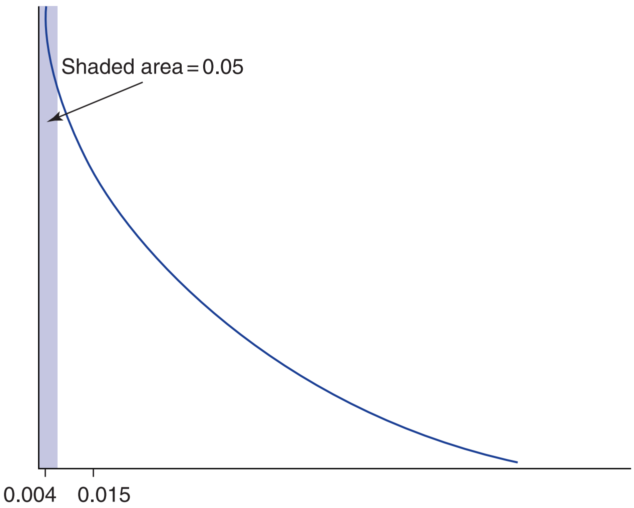

FIGURE 13.4

$\chi^2_{1,0.95} = 0.004$ *implies that* $P\left\{\chi^2_1 \leq 0.015\right\} \geq 0.05$.

Table 13.2 Chi-Squared Goodness-of-Fit Test

Suppose that each member of a population has one of the values $1, 2, \ldots, k$. Let P_i be the proportion of the population that has value i, $i = 1, \ldots, k$. Let p_i, $i = 1, \ldots, k$, be a specified set of nonnegative numbers that sum to 1, and consider a test of

$$\mathrm{H}_0\colon P_i = p_i \qquad \text{for all } i = 1, \ldots, k$$

against

$$\mathrm{H}_1\colon P_i \neq p_i \qquad \text{for some } i = 1, \ldots, k$$

To test this, draw a random sample of n members of the population. Let $e_i = np_i$, $i = 1, \ldots, k$, and make n large enough so that all the e_i are at least 1 and at least 80 percent of them are at least 5.

Let N_i equal the number of members of the sample that have value i. Use the test statistic

$$\mathrm{TS} = \sum_{i=1}^{k} \frac{(N_i - e_i)^2}{e_i}$$

The significance-level-α test is to

Reject H_0	if $\mathrm{TS} \geq \chi^2_{k-1,\alpha}$
Not reject H_0	otherwise

Equivalently, if the value of TS is v, then the p value is given by

$$p \text{ value} = P\{\chi^2_{k-1} \geq v\}$$

Here χ^2_{k-1} is a chi-squared random variable with $k-1$ degrees of freedom and $\chi^2_{k-1,\alpha}$ is the $100(1-\alpha)$th percentile of this distribution.

PROBLEMS

1. Determine the following chi-squared percentile values.
   (a) $\chi^2_{5,\,0.01}$
   (b) $\chi^2_{5,\,0.05}$
   (c) $\chi^2_{10,\,0.01}$
   (d) $\chi^2_{10,\,0.05}$
   (e) $\chi^2_{20,\,0.05}$

2. Consider a data set of 200 elements having the following frequency table:

Outcome	Frequency
1	44
2	38
3	57
4	61

Consider a test of the hypothesis that each of the 200 data values is equally likely to be any of the values 1 through 4.

(a) Express notationally the null and the alternative hypotheses.
(b) Compute the value of the test statistic.
(c) What conclusion is drawn at the 10 percent level of significance?
(d) Repeat part (c), using the 5 percent significance level.
(e) Repeat part (c), using the 1 percent significance level.

(Science Photo Library/ Photo Researchers)

Karl Pearson

Historical Perspective

An epochal event in the history of statistics occurred in 1900, when Karl Pearson published a paper in *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. In this paper he presented his chi-squared goodness-of-fit test. This was an event of great importance because it changed the way people viewed the subject of statistics. Whereas up to then most scientists thought of statistics as a discipline of data organization and presentation, many were led by this paper to view statistics as a discipline concerned with the testing of hypotheses.

3. In a certain county, it has been historically accepted that 52 percent of the patients who go to hospital emergency rooms are in stable condition, 32 percent are in serious condition, and 16 percent are in critical condition. However, a particular county hospital feels that its percentages are different. To prove its claim, the hospital has randomly selected a sample of 300 patients who have visited its emergency room in the past 6 months. The numbers falling in each grouping are as follows:

Stable	148
Serious	92
Critical	60

Do these data prove the claim of the hospital? Explain carefully what the null hypothesis is, and use the 5 percent level of significance.

4. A random sample of 100 student absences yielded the following data on the days of the week on which the absences occurred:

Day	Monday	Tuesday	Wednesday	Thursday	Friday
Frequency	27	19	13	15	26

Test the hypothesis that an absence is equally likely to occur on any of the five days. What are your conclusions?

5. Consider an experiment having six possible outcomes whose probabilities are hypothesized to be 0.1, 0.1, 0.05, 0.4, 0.2, and 0.15. This is to be tested by performing 60 independent replications of the experiment. If the resultant number of times that each of the six outcomes occur is 4, 3, 7, 17, 16, and 13, should the hypothesis be rejected? Use the 5 percent level of significance.

6. In a certain region, 84 percent of drivers have no accidents in a year, 14 percent have exactly one accident, and 2 percent have two or more accidents. In a random sample of 400 lawyers, 308 had no accidents, 66 had one accident, and 26 had two or more. Could you conclude from this that lawyers do not exhibit the same accident profile as the rest of the drivers in the region?

7. The past output of a machine indicates that each unit it produces will be

Top grade	with probability 0.38
High grade	with probability 0.32
Medium grade	with probability 0.26
Low grade	with probability 0.04

A new machine, designed to perform the same job, has produced 500 items with the following results:

Top grade	222
High grade	171
Medium grade	98
Low grade	9

Can the difference in output be ascribed solely to chance? Explain!

8. Roll a die 100 times, keeping track of the frequency of each of the six possible outcomes. Use the resulting data to test the hypothesis that all six sides are equally likely to come up.

9. A marketing manager claims that mail-order sales are equally likely to come from each of four different regions. An employee does not agree and so has collected a random sample of 400 recent orders. They yielded the following numbers from each region:

Region 1	106
Region 2	138
Region 3	76
Region 4	80

Do the data disprove the manager's claim at the 5 percent level of significance? What about at the 1 percent level of significance?

10. A study was instigated to see if southern California earthquakes of at least moderate size (having values of at least 4.4 on the Richter scale) are more likely to occur on certain days of the week than on others. The catalogs yielded the following data on 1100 earthquakes:

Day	Sun	Mon	Tues	Wed	Thurs	Fri	Sat
Number of earthquakes	156	144	170	158	172	148	152

Test, at the 5 percent level, the hypothesis that an earthquake is equally likely to occur on any of the seven days of the week.

11. In certain state lotteries one buys a ticket and then chooses four different integers, between 0 and 36 inclusive. The lottery commission then randomly selects four numbers in this range in such a way that all possible choices are equally likely. After taking out its percentage—sometimes as high as 40 percent—the lottery commission then divides the remainder equally among all those players who had the correct choice of four numbers. Since all possible four-number choices are equally likely to be chosen by the commission, it is easy to see that it is best to select "unpopular" numbers so that if you do win, you will not have to share the prize with too many others. That is, since your chances of being a winner and the amount of money returned to the winners' pool do not depend on your selection, it is best to choose numbers that others are unlikely to choose.

This reasoning raises the question of whether there are indeed unpopular numbers, that is, numbers that are played less frequently than others. To answer this question, one could perform a chi-squared test of the hypothesis that all choices of lottery players are equally likely.

Consider a simplified lottery in which each player selects one of the integers between 1 and 10. Suppose that a random sample of

10,000 previously purchased lottery tickets yielded that each of the 10 numbers had been played with the following frequencies:

Number	Frequency
1	1122
2	1025
3	1247
4	818
5	1043
6	827
7	1149
8	946
9	801
10	1022

Do these data prove that the 10 numbers are not being played with equal frequency?

12. Data provided by the U.S. Bureau of Labor Statistics indicate that the age breakdown, by percentage, of all U.S. workers who are on flexible schedules is as follows:

Age range	Percentage
16–24	13.7
25–34	32.5
35–44	26.3
45–54	17.1
55 and up	10.4

Suppose that a random sample of 240 workers on flexible schedules in the city of Sacramento yielded 24 in the age range of 16 to 24, 94 in the age range of 25 to 34, 48 in the age range of 35 to 44, 35 in the age range of 45 to 54, and 39 in the age range of 55 and up. Can we conclude, at the 5 percent level of significance, that the age breakdown of Sacramento workers on flexible schedules is different from the national breakdown?

13. The 1995 annual report of the Girl Scouts of America indicates that 59.8 percent of members were 8 years old or younger, 32.4 percent were between 9 and 11 years old, and 7.8 percent were 12 years and older. In 2002, a random sample of 400 Girl Scouts contained 255 who were 8 years old or younger, 112 who were between 9 and 11 years old, and 33 who were 12 years old or over. Test the hypothesis that the 2002 percentages are the same as they were in 1995. Use the 5 percent level of significance.

14. Karl Pearson reported that he flipped a coin 24,000 times, with 12,012 heads and 11,988 tails resulting. Is this believable? Explain the reasoning behind your answer.
15. The following gives the age breakdown, by percentages, of unmarried women having children in 1986:

Age range	Percentage
14 or less	1.1
15–19	32.0
20–24	36.0
25–29	18.9
30 and up	12.0

Source: U.S. National Center for Health Statistics, *Vital Statistics of the United States.*

A recent random sample of 1000 births to unmarried women indicated that 42 of the mothers were age 14 or younger, 403 were between 15 and 19 years old, 315 were between 20 and 24 years old, 150 were between 25 and 29 years old, and 90 were 30 years or older. Do these data prove that today's percentages differ from those in 1986?

13.3 TESTING FOR INDEPENDENCE IN POPULATIONS CLASSIFIED ACCORDING TO TWO CHARACTERISTICS

Consider a large population in which each member is classified according to two distinct characteristics, which we shall designate as the X characteristic and the Y characteristic. Suppose that the possible values for the X characteristic are denoted as 1 or 2 or . . . or r; similarly, the possible values of the Y characteristic are denoted as 1 or 2 or . . . or s. Thus, there are r possible values for the X characteristic and s possible values for the Y characteristic.

■ Example 13.5

Consider a population of voting-age adults, and suppose that each adult is classified according to both gender—female or male—and political affiliation—Democrat, Republican, or Independent. Let the X characteristic represent gender and the Y characteristic represent political affiliation. Since there are two possible genders and three possible political affiliations, $r = 2$ and $s = 3$. Let us say that a person's X characteristic is 1 if the person is a woman and 2 if the person is a man. Also, say that a person's Y characteristic is 1 if the person is a

Democrat, 2 if the person is a Republican, and 3 if he or she is an Independent. Thus, for instance, a woman who is a Republican would have X characteristic 1 and Y characteristic 2. ■

Let P_{ij} denote the proportion of the population that has both X characterization i and Y characterization j, for i being any of the values $1, 2, \ldots, r$ and j being any of the values $1, 2, \ldots, s$. Also, let P_i denote the proportion of the population who have X characteristic i, and let Q_j be the proportion who have Y characteristic j. Thus if X and Y denote the values of the X characteristic and Y characteristic of a randomly chosen member of the population, then

$$P\{X = i, Y = j\} = P_{ij}$$
$$P\{X = i\} = P_i$$
$$P\{Y = j\} = Q_j$$

■ Example 13.6

For the situation described in Example 13.5, P_{11} represents the proportion of the population consisting of women who classify themselves as Democrats, P_{12} is the proportion of the population consisting of women who classify themselves as Republicans, and P_{13} is the proportion of the population consisting of women who classify themselves as Independents. The proportions P_{21}, P_{22}, and P_{23} are defined similarly, with *men* replacing *women* in the definitions. The quantities P_1 and P_2 are the proportions of the population that are, respectively, women and men; Q_1, Q_2, and Q_3 are the proportions of the population that are, respectively, Democrats, Republicans, and Independents. ■

We will be interested in developing a test of the hypothesis that the X characteristic and Y characteristic of a randomly chosen member of the population are independent. Recalling that X and Y are independent if

$$P\{X = i, Y = j\} = P\{X = i\}P\{Y = j\}$$

it follows that we want to test the null hypothesis

$$\mathrm{H}_0\colon P_{ij} = P_i Q_j \quad \text{for all } i = 1, \ldots, r,\ j = 1, \ldots, s$$

against the alternative

$$\mathrm{H}_1\colon P_{ij} \neq P_i Q_j \quad \text{for some values of } i \text{ and } j$$

To test this hypothesis of independence, we start by choosing a random sample of size n of members of the population. Let N_{ij} denote the number of elements of the sample that have both X characteristic i and Y characteristic j.

■ Example 13.7

Consider Example 13.5, and suppose that a random sample of 300 people were chosen from the population, with the following data resulting:

	j			
i	**Democrat**	**Republican**	**Independent**	**Total**
Women	68	56	32	156
Men	52	72	20	144
Total	120	128	52	300

Thus, for instance, the random sample of size 300 contained 68 women who classified themselves as Democrats, 56 women who classified themselves as Republicans, and 32 women who classified themselves as Independents; that is, $N_{11} = 68$, $N_{12} = 56$, and $N_{13} = 32$. Similarly, $N_{21} = 52$, $N_{22} = 72$, and $N_{23} = 20$.

This table, which specifies the number of members of the sample that fall in each of the rs cells, is called a *contingency table*. ■

If the hypothesis is true that the X and Y characteristics of a randomly chosen member of the population are independent, then each element of the sample will have X characteristic i and Y characteristic j with probability P_iQ_j. Hence, if these probabilities were known then, from the results of Sec. 13.2, we could test H_0 by using the test statistic

$$\text{TS} = \sum_i \sum_j \frac{(N_{ij} - e_{ij})^2}{e_{ij}}$$

where

$$e_{ij} = nP_iQ_j$$

The quantity e_{ij} represents the expected number, when H_0 is true, of elements in the sample that have both X characteristic i and Y characteristic j. In computing TS we must calculate the sum of the terms for all rs possible values of the pair i, j. When H_0 is true, TS will have an approximately chi-squared distribution with $rs - 1$ degrees of freedom.

The trouble with using this approach directly is that the $r + s$ quantities P_i and Q_j, $i = 1, \ldots, r$, $j = 1, \ldots, s$, are not specified by the null hypothesis. Thus, we need first to estimate them. To do so, let N_i and M_j denote the number of elements of the sample that have, respectively, X characteristic i and Y characteristic j. Because N_i/n and M_j/n are the proportions of the sample having, respectively, X characteristic i and Y characteristic j, it is natural to use them as estimators of P_i and Q_j.

That is, we estimate P_i and Q_j by

$$\hat{P}_i = \frac{N_i}{n} \qquad \hat{Q}_j = \frac{M_j}{n}$$

This leads to the following estimate of e_{ij}:

$$\hat{e}_{ij} = n\hat{P}_i\hat{Q}_j = \frac{N_i M_j}{n}$$

In words, $\hat{e}_{ij}$ is equal to the product of the number of members of the sample that have X characteristic i (that is, the sum of row i of the contingency table) and the number of members of the sample that have Y characteristic j (that is, the sum of column j of the contingency table) divided by the sample size n.

Thus, it seems that a reasonable test statistic to use in testing the independence of the X characteristic and the Y characteristic is the following:

$$\mathrm{TS} = \sum_i \sum_j \frac{(N_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}}$$

where $\hat{e}_{ij}, i = 1, \ldots, r, j = 1, \ldots, s$, are as just given.

To specify the set of values of TS that will result in rejection of the null hypothesis, we need to know the distribution of TS when the null hypothesis is true. It can be shown that when H_0 is true, the distribution of the test statistic TS is approximately a chi-squared distribution with $(r-1)(s-1)$ degrees of freedom. From this, it follows that the significance-level-α test of H_0 is as follows:

Reject H_0	if $\mathrm{TS} \geq \chi^2_{(r-1)(s-1),\alpha}$
Do not reject H_0	otherwise

A technical remark: It is not difficult to see why the test statistic TS should have $(r-1)(s-1)$ degrees of freedom. Recall from Sec. 13.2 that if all the values P_i and Q_j are specified in advance, then the test statistic has $rs - 1$ degrees of freedom. (This is so since k, the number of different types of elements in the population, is equal to rs.) Now, at first glance it may seem that we have to use the data to estimate $r + s$ parameters. However, since the P_i's and the Q_j's both sum to 1—that is, $\Sigma_i P_i = \Sigma_j Q_j = 1$—we really only need to estimate $r - 1$ of the P_i's and $s - 1$ of the Q_j's. (For instance, if r is equal to 2, then an estimate of P_1 will automatically provide an estimate of P_2 since $P_2 = 1 - P_1$.) Hence, we actually need to estimate $r - 1 + s - 1 = r + s - 2$ parameters. Since a degree of freedom is lost for each parameter estimated, it follows that the resulting test statistic has $rs - 1 - (r + s - 2) = rs - r - s + 1 = (r-1)(s-1)$ degrees of freedom.

■ Example 13.8

The data of Example 13.7 are as follows:

i	j: 1	2	3	Total $= N_i$
1	68	56	32	156
2	52	72	20	144
Total $= M_j$	120	128	52	300

What conclusion can be drawn? Use the 5 percent level of significance.

Solution

From the given data, the six values of

$$\hat{e}_{ij} = \frac{N_i M_j}{n}$$

are as follows:

$$\hat{e}_{11} = \frac{N_1 M_1}{n} = \frac{156 \times 120}{300} = 62.40$$

$$\hat{e}_{12} = \frac{N_1 M_2}{n} = \frac{156 \times 128}{300} = 66.56$$

$$\hat{e}_{13} = \frac{N_1 M_3}{n} = \frac{156 \times 52}{300} = 27.04$$

$$\hat{e}_{21} = \frac{N_2 M_1}{n} = \frac{144 \times 120}{300} = 57.60$$

$$\hat{e}_{22} = \frac{N_2 M_2}{n} = \frac{144 \times 128}{300} = 61.44$$

$$\hat{e}_{23} = \frac{N_2 M_3}{n} = \frac{144 \times 52}{300} = 24.96$$

The value of the test statistic is thus

$$\text{TS} = \frac{(68 - 62.40)^2}{62.40} + \frac{(56 - 66.56)^2}{66.56} + \frac{(32 - 27.04)^2}{27.04} + \frac{(52 - 57.60)^2}{57.60} + \frac{(72 - 61.44)^2}{61.44} + \frac{(20 - 24.96)^2}{24.96} = 6.433$$

Since $r = 2$ and $s = 3$, $(r - 1)(s - 1) = 2$ and so we must compare the value of TS with the critical value $\chi^2_{2,\,0.05}$. From Table 13.1,

$$\chi^2_{2,\,0.05} = 5.991$$

Since TS $\geq$ 5.991, the null hypothesis is rejected at the 5 percent level of significance. That is, the hypothesis that gender and political affiliation of members of the population are independent is rejected at the 5 percent level of significance. ■

The test of the hypothesis that the X and Y characteristics of a randomly chosen member of the population are independent can also be performed by determining the p value of the data. This is accomplished by first calculating the value of the test statistic TS. If its value is v, then the p value is given by

$$p \text{ value} = P\left\{\chi^2_{(r-1)(s-1)} \geq v\right\}$$

where $\chi^2_{(r-1)(s-1)}$ is a chi-squared random variable with $(r-1)(s-1)$ degrees of freedom.

Program 13-2 will calculate the value of the test statistic and then determine the resulting p value. The program supposes that the data are arranged in the form of a contingency table and asks the user to input the successive rows of this table.

■ Example 13.9

A public health scientist wanted to learn about the relationship between the marital status of patients being treated for depression and the severity of their conditions. The scientist chose a random sample of 159 patients who had been treated for depression at a mental health clinic and had these patients classified according to the severity of their depression—severe, normal, or mild—and according to their marital status. The following data resulted.

	Marital status			
Depressive state	**Married**	**Single**	**Widowed or divorced**	**Totals**
Severe	22	16	19	57
Normal	33	29	14	76
Mild	14	9	3	26
Totals	69	54	36	159

Determine the p value of the test of the hypothesis that the depressive state of the clinic's patients is independent of their marital status.

Solution

We run Program 13-2 to obtain that the value of the test statistic and the resulting p value are

$$\text{TS} = 6.828 \qquad p \text{ value} = 0.145$$ ■

The test for independence is summarized in Table 13.3.

Table 13.3 Testing Independence of Two Characteristics of Members of a Population

Assume that each member of a population has both an X and a Y characteristic. Let r and s denote the number of possible X and Y characteristics, respectively. To test

$$\mathrm{H}_0\text{: characteristics of a randomly chosen member are independent}$$

against

$$\mathrm{H}_1\text{: characteristics of a randomly chosen member are not independent}$$

choose a sample of n members of the population. Let N_{ij} denote the number of these that have both X characteristic i and Y characteristic j. Also let

$$N_i = \sum_j N_{ij} \quad \text{and} \quad M_j = \sum_i N_{ij}$$

denote, respectively, the number of members of the sample that have X characteristic i and that have Y characteristic j. The test statistic is

$$\mathrm{TS} = \sum_i \sum_j \frac{(N_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}}$$

where $\hat{e}_{ij} = N_i M_j / n$. The significance-level-α test is as follows:

$$\begin{array}{ll} \text{Reject } \mathrm{H}_0 & \text{if } \mathrm{TS} \geq \chi^2_{(r-1)(s-1),\,\alpha} \\ \text{Do not reject } \mathrm{H}_0 & \text{otherwise} \end{array}$$

Equivalently, if the value of TS is v, then the p value is

$$p \text{ value} = P\{\chi^2_{(r-1)(s-1)} \geq v\}$$

PROBLEMS

1. The following contingency table presents data from a sample of a population that is characterized in two different ways:

	X characteristic		
***Y* characteristic**	**A**	**B**	**C**
1	32	12	40
2	56	48	60

(a) Determine the value of the test statistic when testing that the two characteristics are independent.
(b) Would the null hypothesis be rejected at the 5 percent level of significance?
(c) What about at the 1 percent level?

2. There is some evidence that ownership of a dog may have predictive value in determining whether an individual will survive a heart attack. The following data are from a random sample of 95 individuals who each suffered a severe heart attack. The data classify each of these individuals with respect to (1) whether they were still alive 1 year after their attack and (2) whether they had a dog as a pet.

	Had pet	No pet
Survived	28	44
Did not survive	8	15

Do these data prove, at the 5 percent level of significance, that owning a pet and survival are dependent? Explain carefully what null hypothesis you are testing and what test statistic you are using.

Historical Perspective

Her biographer called her the *passionate statistician.* It was an appropriate description of Florence Nightingale, the woman who almost single-handedly changed nursing into a science. During the Crimean war she searched out and collected data on sanitary conditions and mortality rates in military hospitals, and she used these data to statistically prove that the two were dependent. Her work was instrumental in improving the hygienic conditions in hospitals and resulted in the saving of untold lives.


(Photoworld/FPG)

Florence Nightingale

Florence Nightingale was a follower of the Belgian statistician Adolphe Quetelet, and she believed, as he did, that "accidents occur with astonishing regularity when the same conditions exist." She held that successful administrators were ones who searched out data. She felt that the universe evolved in accordance with a divine plan and that it was every person's job to learn to live in harmony with it. But to understand this plan, she believed that one had to study statistics. In the words of Karl Pearson, "For Florence Nightingale, statistics were more than a study, they were her religion."

3. A random sample of 187 voters were chosen, and the voters were asked to evaluate the performance of the first 100 days of the U.S. President. Use the resulting data to test the hypothesis that the evaluation

of an individual does not depend on whether that individual is a man or woman.

	Women	Men
Positive evaluation	54	47
Negative evaluation	20	32
Not sure	23	11

Use the 5 percent level of significance.

4. An insurance company is interested in determining whether there is a relationship between automobile accident frequency and cigarette smoking. It randomly sampled 597 policyholders and came up with the following data:

Number of accidents in last 2 years	Smokers	Nonsmokers
0	35	170
1	79	190
2 or more	57	66

Test the hypothesis, at the 5 percent level of significance, that the accident frequency of a randomly chosen policyholder is independent of his or her smoking habits.

5. The management of a certain hotel is interested in whether all its guests are treated the same regardless of the prices of their rooms. They randomly chose 155 recent guests and questioned them about the service they had received at the hotel. The following summary data resulted:

	Type of room		
Service ranking	**Economy**	**Standard**	**Luxury**
Excellent	30	21	9
Good	36	29	8
Fair	12	8	2

What conclusions would you draw?

6. The following data categorize a random selection of professors of a certain university according to their teaching performance (as measured by the students in their classes) in the most recent semester and the number of courses they were teaching at the time.

	Number of courses		
Student ranking	**1**	**2**	**3 or more**
Above average	12	10	4
Average	32	40	38
Below average	7	12	25

Test, at the 5 percent level, the hypothesis that a professor's teaching performance is independent of the number of courses she or he is teaching.

7. In Prob. 6, it is possible that only certain professors, usually ones who specialize in research, would teach only one course in a semester. These would then tend to be more advanced courses and to have fewer students than in most courses. Thus, to learn more directly whether teaching additional courses affects teaching performance, it might be reasonable to consider the data of Prob. 6 with the column pertaining to those teaching only one class deleted. Make this change and repeat Prob. 6.

8. The socioeconomic status of residents of a particular neighborhood can be classified as either lower or middle class. A sample of residents were questioned about their attitude toward a planned public health clinic for the neighborhood. The results are as follows:

	Socioeconomic class	
Attitude	**Lower**	**Middle**
In favor	87	63
Against	46	55

Test the hypothesis, at the 5 percent level of significance, that lower- and middle-class residents of the neighborhood have the same attitude toward the new clinic.

9. A market research firm has distributed samples of a new shampoo to a variety of individuals. The following data summarize the comments of these individuals about the shampoo as well as provide the age group into which they fall.

	Age group (years)		
Rating	**15–20**	**21–30**	**Over 30**
Excellent	18	20	41
Good	25	27	43
Fair	17	15	26
Poor	3	2	8

Do these data prove that different age groups have different opinions about the shampoo? Use the 5 percent level of significance.

10. A random sample of 160 patients at a health maintenance organization yielded the following information about their smoking status and blood cholesterol counts:

	Blood cholesterol count		
Smoking status	**Low**	**Moderate**	**High**
Heavy	6	14	24
Light	12	23	15
Nonsmoker	23	32	11

(a) Would the hypothesis of independence between blood cholesterol count and smoking status be rejected at the 5 percent level of significance?

(b) Repeat part (a), but this time use the 1 percent level of significance.

(c) Do your results imply that a reduction in smoking will result in a lowered blood cholesterol level? Explain!

11. To see if there was any dependency between the type of professional job held and one's religious affiliation, a random sample of 638 individuals belonging to a national organization of doctors, lawyers, and engineers were chosen in a study. The results of the sample are given in the following contingency table:

	Doctors	Lawyers	Engineers
Protestant	64	110	152
Catholic	60	86	78
Jewish	57	21	10

Test the hypothesis, at the 5 percent level of significance, that the profession of individuals in this organization and their religious affiliation are independent. Repeat at the 1 percent level.

12. Look at the following contingency table and guess (without any computations) as to the result of a test, at the 5 percent level of significance, of the hypothesis that the two data characteristics are independent.

	A	B	C
1	26	44	30
2	14	30	25
3	30	45	33

Now perform the computations.

13. Repeat Prob. 11, but first double all the data values.
14. Repeat Prob. 12, but first double all the data values.

13.4 TESTING FOR INDEPENDENCE IN CONTINGENCY TABLES WITH FIXED MARGINAL TOTALS

In Example 13.5 we were interested in determining whether gender and political affiliation were dependent in a particular population. To test this hypothesis, we first chose a random sample of people from this population and then noted their characteristics. However, another way in which we could gather data is to fix in advance the numbers of men and women in the sample and then to choose random samples of those sizes from the subpopulations of men and women. That is, rather than let the numbers of women and men in the sample be determined by chance, we might decide these numbers in advance. Because doing so would result in fixed specified values for the total numbers of men and women in the sample, the resulting contingency table is often said to have *fixed margins* (since the totals are given in the margins of the table).

It turns out that even when the data are collected in the manner just prescribed, the same hypothesis test as given in Sec. 13.3 can be used to test for the independence of the two characteristics. The test statistic remains

$$\mathrm{TS} = \sum_i \sum_j \frac{(N_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}}$$

where

N_{ij} = number of members of sample who have both X characteristic i and Y characteristic j

N_i = number of members of sample who have X characteristic i

M_j = number of members of sample who have Y characteristic j

and

$$\hat{e}_{ij} = \frac{N_i M_j}{n}$$

where n is the total size of the sample.

In addition, it is still true that when H_0 is true, TS will have an approximately chi-squared distribution with $(r-1)(s-1)$ degrees of freedom. (The quantities r and s refer, of course, to the numbers of possible values of the X and Y characteristic, respectively.) In other words, the test of the independence hypothesis is unaffected by whether the marginal totals of one characteristic are fixed in advance or result from a random sample of the entire population.

■ Example 13.10

A randomly chosen group of 20,000 nonsmokers and one of 10,000 smokers were followed over a 10-year period. The following data relate the numbers developing lung cancer in that period.

	Smokers	Nonsmokers	Total
Lung cancer	62	14	76
No lung cancer	9,938	19,986	29,924
Total	10,000	20,000	30,000

Test the hypothesis that smoking and lung cancer are independent. Use the 1 percent level of significance.

Solution

The estimated numbers expected to fall in each *ij* cell are

$$\hat{e}_{11} = \frac{(76)(10{,}000)}{30{,}000} = 25.33$$

$$\hat{e}_{12} = \frac{(76)(20{,}000)}{30{,}000} = 50.67$$

$$\hat{e}_{21} = \frac{(29{,}924)(10{,}000)}{30{,}000} = 9974.67$$

$$\hat{e}_{22} = \frac{(29{,}924)(20{,}000)}{30{,}000} = 19{,}949.33$$

Therefore, the value of the test statistic is

$$\begin{aligned}\text{TS} &= \frac{(62-25.33)^2}{25.33} + \frac{(14-50.67)^2}{50.67} + \frac{(9938-9974.67)^2}{9974.67} \\ &\quad + \frac{(19{,}986-19{,}949.33)^2}{19{,}949.33} \\ &= 53.09 + 26.54 + 0.13 + 0.07 = 79.83\end{aligned}$$

Since this is far larger than $\chi^2_{1,0.01} = 6.635$, we reject the null hypothesis that whether a randomly chosen person develops lung cancer is independent of whether that person is a smoker. ■

We now show how we can use the framework of this section to test the hypothesis that m population proportions are equal. To begin, consider m separate populations of individuals. Suppose that the proportion of members of the ith population that are in favor of a certain proposition is p_i, and consider a test of the null hypothesis that all the p_i's are equal. That is, consider a test of

$$\mathrm{H}_0\colon p_1 = p_2 = \cdots = p_m$$

against

$$\mathrm{H}_1\colon \text{not all the } p_i\text{'s are equal}$$

To obtain a test of this null hypothesis, consider first the superpopulation consisting of all members of each of the m populations. Any member of this superpopulation can be classified according to two characteristics. The first characteristic specifies which of the m populations the member is from, and the second characteristic specifies whether the member is in favor of the proposition. Now, the hypothesis that all the p_i's are equal is just the hypothesis that the same proportions of members of each population are in favor of the proposition. But this is exactly the same as stating that for members of the superpopulation the characteristic of being for or against the proposition is independent of the population that the member is from. That is, the null hypothesis H_0 is equivalent to the hypothesis of independence of the two characteristics of the superpopulation.

Therefore, we can test H_0 by first choosing independent random samples of fixed sizes from each of the m populations. If we let M_i be the sample size from population i, for $i = 1, \ldots, m$, we can test H_0 by testing for independence in the following contingency table:

	Population				
	1	**2**	**...**	m	**Total**
In favor	F_1	F_2	$\cdots$	F_m	N_1
Against	A_1	A_2	$\cdots$	A_m	N_2
Total	M_1	M_2	$\cdots$	M_m	

Here, F_i refers to the number of members of population i who are in favor, and A_i refers to the number who are against the proposition.

■ Example 13.11

A recent study reported that 500 female office workers were randomly chosen and questioned in each of four different countries. One of the questions related to whether these women often received verbal or sexual abuse on the job. The following data resulted:

Country	Number reporting abuse
Australia	28
Germany	30
Japan	58
United States	55

Based on these data, is it plausible that the proportions of female office workers who often feel abused at work are the same for these countries?

Solution

Putting the data in the form of a contingency table gives the following:

	Country				
	1	2	3	4	Total
Receive abuse	28	30	58	55	171
Do not receive abuse	472	470	442	445	1829
Total	500	500	500	500	2000

We can now test the null hypothesis by testing for independence in the preceding contingency table. If we run Program 13-2, then the value of the test statistic and the resulting p value are

$$\text{TS} = 19.51 \qquad p \text{ value} = 0.0002$$

Therefore, the hypothesis that the percentages of women who feel they are being abused on the job are the same for these countries is rejected at the 1 percent level of significance (and, indeed, at any significance level above 0.02 percent). ■

When there are only two populations, the preceding test of the equality of population proportions is identical to the one presented in Sec. 10.6.

PROBLEMS

1. Can we conclude from the results of Example 13.10 that smoking causes lung cancer? What other explanations are possible?
2. A study of the relationship between school preferences and family income questioned 100 upper-income and 100 lower-income families

in a certain city about the type of school they would most like their children to attend. These are the resulting data:

Preference	Upper income	Lower income
Public	22	19
Private religious	31	39
Private nonreligious	47	42

What conclusions can you draw?

3. A sample of 300 cars having cellular phones and one of 400 cars without phones were tracked for 1 year. The following table gives the number of these cars involved in accidents over that year.

	Accident	No accident
Cellular phone	22	278
No Phone	26	374

Use these data to test the hypothesis that having a cellular phone in your car and being involved in an accident are independent. Use the 5 percent level of significance.

4. A newspaper chain sampled 100 readers of each of its three major newspapers to determine their economic class. The results were as follows:

	Newspaper		
Economic class	1	2	3
Lower middle	22	25	28
Middle	41	37	44
Upper middle	37	38	28

Test the hypothesis that the newspaper an individual reads and the economic class to which that person belongs are independent. Use the 5 percent level.

5. The following table shows the number of defective and acceptable items in samples taken both before and after the introduction of a modification in the manufacturing process.

	Defective	Nondefective
Before	22	404
After	18	422

Does the data prove that the modification results in a different percentage of defective items?

6. From a statistics class of 200 students, 100 were randomly chosen to watch the lectures on television rather than in person. The other 100 stayed in the lecture hall. The final grades of the students are as follows:

	A	B	C	Less than C
In-class students	22	38	35	5
Television students	18	32	40	10

Test the hypothesis that final grades are independent of whether the student watches on television or is present in the lecture hall. Can we reject at the 5 percent level of significance? What about at the 1 percent level?

7. To study the effect of fluoridated water supplies on tooth decay, two communities of roughly the same socioeconomic status were chosen. One of these communities had fluoridated water, while the other did not. Random samples of 200 teenagers from both communities were chosen, and the numbers of cavities they had were determined. The following data resulted:

Cavities	Fluoridated town	Nonfluoridated town
0	154	133
1	20	18
2	14	21
3 or more	12	28

Do these data establish, at the 5 percent level of significance, that the number of dental cavities a person has is not independent of whether that person's water supply is fluoridated? What about at the 1 percent level?

8. An automobile dealership sent out postcards to 990 potential customers, offering them a free test drive of one of its cars. Each postcard was colored red, white, light blue, or green. Here are data relating the number of customers who responded and the colors of the postcards they had been sent:

	Red	White	Blue	Green
Responded	108	106	105	127
No response	142	144	135	123

Test the hypothesis that the color of the postcard sent does not affect the recipient's chance of responding. Use the 5 percent level of significance.

9. Random samples of 50 college students, 40 college faculty, and 60 bankers yielded the following data relating to the numbers of smokers in these samples:

Group	Number who smoke
College students	18
College faculty	12
Bankers	24

(a) Test the hypothesis, at the 10 percent level of significance, that the same percentages of college students, college faculty, and bankers are smokers.
(b) Repeat part (a) at the 5 percent level of significance.
(c) Repeat part (a) at the 1 percent level of significance.

10. To determine if a malpractice lawsuit is more likely to follow certain types of surgery, random samples of three different types of surgeries were studied, and the following data resulted:

Type of operation	Number sampled	Number leading to a lawsuit
Heart surgery	400	16
Brain surgery	300	19
Appendectomy	300	7

Test the hypothesis that the percentages of the surgical operations that lead to lawsuits are the same for each of the three types.
(a) Use the 5 percent level of significance.
(b) Use the 1 percent level of significance.

KEY TERMS

Goodness-of-fit test: A statistical test of the hypothesis that a specified set of k probabilities represents the proportion of members of a large population that fall into each of k distinct categories.

Contingency table: A table that classifies each element of a sample according to two distinct characteristics.

SUMMARY

Goodness-of-Fit Tests Consider a large population of elements, each of which has a value that is 1 or 2 or ... or k. Let P_i denote the proportion of the population that has value i, for $i = 1, \ldots, k$. For a given set of probabilities $p_1, \ldots, p_k (p_i \geq 0, \sum_i p_i = 1)$, consider a test of

$$H_0: P_i = p_i \quad \text{for all } i = 1, \ldots, k$$

against the alternative

$$H_1: P_i \neq p_i \quad \text{for some } i, i = 1, \ldots, k$$

To test this null hypothesis, first draw a random sample of n elements of the population. Let N_i denote the number of elements in the sample that have value i. The test statistic to be employed is

$$TS = \sum_{i=1}^{k} \frac{(N_i - e_i)^2}{e_i}$$

where

$$e_i = np_i$$

When H_0 is true, e_i is equal to the expected number of elements in the sample that have value i.

The hypothesis test is to reject H_0 when TS is sufficiently large. To determine how large, we use the fact that when H_0 is true, TS has a distribution that is approximately a chi-squared distribution with $k - 1$ degrees of freedom. This implies that the significance-level-α test is to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } TS \geq \chi^2_{k-1,\alpha} \\ \text{Not reject } H_0 & \text{otherwise} \end{array}$$

The quantity $\chi^2_{k-1,\alpha}$ is defined by

$$P\{\chi^2_{k-1} \geq \chi^2_{k-1,\alpha}\} = \alpha$$

where χ^2_{k-1} is a chi-squared random variable with $k - 1$ degrees of freedom.

This test is called the *chi-squared goodness-of-fit test.* It can also be implemented by determining the p value of the data set. If the observed value of TS is v, then the p value is

$$p \text{ value } = P\{\chi^2_{k-1} \geq v\}$$

Program 13-1 can be used to determine the value of the test statistic and the resulting p value.

Testing for Independence in Populations Whose Elements Are Classified According to Two Different Characteristics Suppose now that each element of a population is classified according to two distinct characteristics, which we call the X characteristic and the Y characteristic. Suppose the possible values of the X characteristic are 1 or 2 or ... or r, and the possible values for the Y characteristic are 1 or 2 or ... or s. Let P_{ij} denote the proportion of the population that has X characteristic i and Y characteristic j. Let P_i denote the proportion of the population whose X characteristic is $i, i = 1, \ldots, r$; and let Q_j denote the proportion whose Y characteristic is $j, j = 1, \ldots, s$.

Consider a test of the null hypothesis that the X and Y characteristics of a randomly chosen member of the population are independent. That is, consider a test of

$$H_0: P_{ij} = P_i Q_j \qquad \text{for all } i, j$$

against

$$H_1: P_{ij} \neq P_i Q_j \qquad \text{for some } i, j$$

To test this, draw a random sample of n elements of the population. Let N_{ij} denote the number of these having X characteristic i and Y characteristic j. Also let N_i denote the number that have X characteristic i; and let M_j denote the number that have Y characteristic j. The test statistic used to test H_0 is

$$\text{TS} = \sum_i \sum_j \frac{(N_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}}$$

where

$$\hat{e}_{ij} = \frac{N_i M_j}{n}$$

The summation in the expression for TS is over all rs possible values of the pair i, j. The significance-level-α test is to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if TS} \geq \chi^2_{(r-1)(s-1),\alpha} \\ \text{Not reject } H_0 & \text{otherwise} \end{array}$$

Equivalently, the test can be implemented by computing the p value. If the value of TS is v, then the p value is given by

$$p \text{ value} = P\{\chi^2_{(r-1)(s-1)} \geq v\}$$

Program 13-2 can be used to determine both the value of TS and the resulting p value.

The same test as just given can also be employed when the sample chosen is not a random sample from the entire population but rather a collection of random samples of fixed sizes from the r (or s) subpopulations whose X characteristic (or Y characteristic) is fixed.

The following tabular presentation of the data is called a *contingency table:*

	Y characteristic						
X characteristic	**1**	**2**	$\cdots$	j	$\cdots$	s	**Total**
1	N_{11}	N_{12}	$\ldots$	N_{1j}	$\ldots$	N_{1s}	N_1
i	N_{i1}	N_{i2}	$\ldots$	N_{ij}	$\ldots$	N_{is}	N_i
r	N_{r1}	N_{r2}	$\ldots$	N_{rj}	$\ldots$	N_{rs}	N_r
Total	M_1	M_2	$\ldots$	M_j	$\ldots$	M_s	n

REVIEW PROBLEMS

1. The following data classify minor accidents over the past year at a certain industrial plant, according to the time periods in which these accidents occurred.

Time period	Number of accidents
8–10 a.m.	47
10–12 p.m.	52
1–3 p.m.	57
3–5 p.m.	63

Test the hypothesis that each accident was equally likely to occur in any of the four time periods. Use the 5 percent level of significance.

2. A movie distribution company often sets up sneak previews of new movies. In such a situation a movie whose title had not previously been announced is shown in addition to the regularly scheduled movie. When the audience leaves the theater, individuals are given questionnaires to be filled out at home and mailed back to the company. Such information is then used by the company in deciding how widely to distribute the movie. Of some interest is whether the popularity of a movie will be the same over different parts of the country. To test this hypothesis, a sneak preview was scheduled in four theaters around the

country. One theater was in New York, one in Chicago, one in Phoenix, and one in Seattle. The following are ratings given to the movie by audiences in these four places:

	Location			
Rating	**New York**	**Chicago**	**Phoenix**	**Seattle**
Excellent	234	141	108	142
Good	303	256	165	170
Poor	102	88	41	45

Test, at the 5 percent level, the hypothesis that the audience reaction is independent of the location. What about at the 1 percent level?

3. Suppose that a die is rolled 600 times. Consider a test of the hypothesis that each roll is equally likely to be any of the six faces. Make up data you think will result in a p value approximately equal to
   (a) 0.50
   (b) 0.05
   (c) 0.95
   (d) Approximate the actual p values for the data you presented in parts (a), (b), and (c).
4. It has been claimed that the proportions of voters presently favoring the Democratic, Republican, or Independent candidate in an upcoming election are 40, 42, and 18 percent. To test this hypothesis, a random sample of 50 voters yielded the following results:

	Democrat	**Republican**	**Independent**
Number favoring	18	22	10

Is the claim consistent with the preceding data? Use the 5 percent level of significance.

5. The following data come from a study of randomly selected automobile accidents. It categorizes each accident by the weight of the car involved and the severity of injury suffered by the driver.

	Weight of car (pounds)		
Injury	**Less than 2500**	**2500–3000**	**Greater than 3000**
Very severe	34	22	8
Average	43	41	47
Moderate	51	60	50

Test, at the 5 percent level of significance, the hypothesis that the severity of injury and the weight of the car are independent.

6. A friend reported the following results when rolling a die 1000 times:

Outcome	Frequency
1	167
2	165
3	167
4	166
5	167
6	168

Do you believe these results? Explain!

7. A random sample of 527 earthquakes in western Japan yielded the following frequencies of occurrence at certain time periods in a day:

Time period	Frequency
12 a.m.–6 a.m.	123
6 a.m.–12 p.m.	135
12 p.m.–6 p.m.	141
6 p.m.–12 a.m.	128

Test the hypothesis that earthquakes are equally likely to occur in each of the four time periods.

8. The following data give the number of murders, by day of the week, in the state of Utah from 1978 through 1990:

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Number	109	74	97	94	83	107	100

Test the hypothesis, at the 5 percent level, that a murder was equally likely to occur on any of the 7 days of the week.

9. The following table gives the cumulative percentage distribution of the heights of U.S. women residents aged 18 to 24 in the years 1976 through 1980:

Height	Women's cumulative percentage
5 feet 0 inches	4.22
5 feet 3 inches	29.06
5 feet 5 inches	58.09
5 feet 7 inches	85.37
5 feet 8 inches	92.30

Source: U.S. National Center for Health Statistics, *Vital and Health Statistics*, series 11, no. 238.

Thus, for instance, 4.22 percent of women in those years were less than or equal to 5 feet 0 inches tall, 24.84 percent were larger than 5 feet 0 inches but less than or equal to 5 feet 3 inches in height, and so on.

Suppose that a random selection of 200 present-day women in the age bracket of 18 to 24 resulted in 6 of them being less than 5 feet 0 inches, 42 of them being between 5 feet 0 inches and 5 feet 3 inches, 48 of them being between 5 feet 3 inches and 5 feet 5 inches, 60 of them being between 5 feet 5 inches and 5 feet 7 inches, 21 being between 5 feet 7 inches and 5 feet 8 inches, and the rest being taller than 5 feet 8 inches. Would these data imply that the height distribution has changed? Use the 5 percent level of significance.

10. One of Mendel's breeding experiments resulted in the following data:

Type of pea	Expected	Observed
Smooth yellow	313	315
Wrinkled yellow	104	101
Smooth green	104	108
Wrinkled green	35	32

Do you think such a good fit is "too good to be true"?

11. A public health clinic detailed the results of 260 elderly patients who had been advised to have a flu vaccine. A total of 184 agreed to have the vaccine, while the other 76 declined. The flu season results for this group were as follows:

	Vaccine	No vaccine
Flu	10	6
No flu	174	70

Does the data establish that those receiving the vaccine had a different chance of contracting the flu from those not receiving the vaccine? Use the 5 percent level of significance. If it does, check at the 1 percent level of significance; if not, check at the 10 percent level.

12. A random sample of 262 married men in their fifties were classified according to their education and number of children. The following contingency table describes the data:

	Number of children		
Education	**0–1**	**2–3**	**More than 3**
Elementary	10	28	22
Secondary	19	63	38
College	14	41	27

Test the hypothesis that the size of a family is independent of the educational level of the father. Use the 5 percent level of significance.

13. The following data relate a mother's age and the birth weight (in grams) of her child:

	Birth weight (grams)	
Mother's age (years)	**Less than 2500**	**More than 2500**
20 or less	12	50
Greater than 20	18	125

(a) Test the hypothesis, at the 5 percent level of significance, that the baby's birth weight is independent of the mother's age.

(b) What is the p value?

14. Repeat Prob. 13 when the four data values are all doubled.

15. On a course evaluation form, students are asked to rank the course as excellent, fair, or poor. In addition, students signify whether the course is required or not for them. A random sample of 121 such evaluations yielded the following data:

	Rating		
	Excellent	**Average**	**Poor**
Required	14	42	18
Not required	12	28	7

Test, at the 5 percent level, the hypothesis that course rating is independent of whether the course is required. What about at the 1 percent level?

16. A class of 154 students in statistics meets in a room that can hold 250 students. Out of curiousity, the instructor categorized each student by gender and seat location. Using the following data, test the hypothesis that these characterizations are independent.

	Front	Middle	Back
Females	22	40	18
Males	10	38	26

17. One might imagine that the first digits of numbers found in an almanac would be equally likely to be 1 or 2 or . . . or 9. Make a random selection of numbers from an almanac, and note the first digit of each. Use the data to test the hypothesis that all nine digits are equally likely.

18. Repeat Prob. 17, this time using the second digit.