tially related to both the row and column classification variables, and must be controlled for. We discuss methods for accomplishing this in Chapter 13. In this chapter, we assume no confounding is present. Thus, we use either the two-sample test for binomial proportions (Equation 10.3) or the equivalent chi-square test for 2×2 contingency tables (Equation 10.5).

SECTION 10.3 Fisher's Exact Test

In Section 10.2, we discussed methods for comparing two binomial proportions using either normal-theory or contingency-table methods. Both methods yield identical p-values. However, they require that the normal approximation to the binomial distribution be valid, which is not always the case, especially for small sample sizes.

Example 10.16 **Cardiovascular Disease, Nutrition** Suppose we wish to investigate the relationship between high salt intake and the occurrence of death from cardiovascular disease (CVD). Groups of high- and low-salt users could be identified and followed over a long period of time to compare the relative frequency of death from CVD in the two groups. In contrast, a much less expensive study would involve looking at death records, separating the CVD deaths from the non-CVD deaths, and then asking a close relative (such as a spouse) about the dietary habits of the deceased, and comparing salt intake of CVD deaths and non-CVD deaths.

The latter type of study, a retrospective study, may be impossible to perform for a number of reasons. But if it is possible, it will almost always be less expensive than the former type of study, a prospective study.

Example 10.17 **Cardiovascular Disease, Nutrition** Suppose a retrospective study is done among men aged 50–54 in a specific county who died over a 1-month period. The investigators attempt to include approximately an equal number of men who died from CVD (the cases) and men who died from other causes (the controls). It is found that of 35 people who died from CVD, 5 were on a high-salt diet before they died, whereas of 25 people who died from other causes, 2 were on such a diet. These data, presented in Table 10.9, are in the form of a 2×2 contingency table, and thus the methods of Section 10.2.2 may be applicable.

TABLE 10.9 **Data concerning the possible relationship between cause of death and high salt intake**

	Type of diet		
Cause of death	High salt	Low salt	Total
Non-CVD	2	23	25
CVD	5	30	35
Total	7	53	60

However, the expected values of this table are too small for such methods to be valid. Indeed,

$$E_{11} = 7(25)/60 = 2.92$$
$$E_{21} = 7(35)/60 = 4.08$$

and thus two of the four cells have expected values less than 5. How should the possible association between cause of death and type of diet be assessed?

In this case, **Fisher's exact test** can be used. This procedure gives exact results for any 2 × 2 table but is only necessary for tables with small expected values, where the standard chi-square test as given in Equation 10.5 is not applicable. For tables in which the use of the chi-square test is appropriate, the two tests give very similar results. Suppose the probability that a man was on a high-salt diet given that his cause of death was noncardiovascular (non-CVD) = p_1, and the probability that a man was on a high-salt diet given that his cause of death was cardiovascular (CVD) = p_2. We wish to test the hypothesis H_0: $p_1 = p_2 = p$ versus H_1: $p_1 \neq p_2$. Table 10.10 gives the general layout of the data.

TABLE 10.10 **General layout of data for Fisher's exact test example**

	Type of diet		
Cause of death	High salt	Low salt	Total
Non-CVD	a	b	$a + b$
CVD	c	d	$c + d$
Total	$a + c$	$b + d$	n

For mathematical convenience, we will assume that the margins of this table are *fixed*; that is, the numbers of non-CVD deaths and CVD deaths are fixed at $a + b$ and $c + d$, respectively, and the numbers of people on high- and low-salt diets are fixed at $a + c$ and $b + d$, respectively. Indeed, it is difficult to compute exact probabilities unless one makes the assumption of fixed margins. The *exact* probability of observing the table with cells a, b, c, d is given as follows:

EQUATION 10.7

Exact Probability of Observing a Table with Cells *a, b, c, d*

$$Pr(a,b,c,d) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!\,a!\,b!\,c!\,d!}$$

The formula in Equation 10.7 is easy to remember, because the numerator is the product of the factorials of each of the row and column margins, and the denominator is the product of the factorial of the grand total and the factorials of the individual cells.

Example 10.18 Suppose we have the 2×2 table as shown in Table 10.11. Compute the exact probability of obtaining this table assuming that the margins are fixed.

Solution

$$Pr(2,5,3,1) = \frac{7!\,4!\,5!\,6!}{11!\,2!\,5!\,3!\,1!} = \frac{5040(24)(120)(720)}{39{,}916{,}800(2)(120)(6)} = \frac{1.0450944 \times 10^{10}}{5.7480192 \times 10^{10}} = .182$$

TABLE 10.11 **Hypothetical 2×2 contingency table in Example 10.18**

2	5	7
3	1	4
5	6	11

10.3.1 The Hypergeometric Distribution

Suppose we consider all possible tables with fixed row margins denoted by N_1 and N_2 and fixed column margins denoted by M_1 and M_2. We assume that the rows and columns have been rearranged so that $M_1 \le M_2$ and $N_1 \le N_2$. We will refer to each table by its (1, 1) cell, because all other cells are then determined from the fixed row and column margins. Let the random variable X denote the cell count in the (1, 1) cell. The probability distribution of X is given by

EQUATION 10.8

$$Pr(X = a) = \frac{N_1!\,N_2!\,M_1!\,M_2!}{N!\,a!(N_1 - a)!(M_1 - a)!(M_2 - N_1 + a)!},\ a = 0, \ldots, \min(M_1, N_1)$$

and $N = N_1 + N_2 = M_1 + M_2$. This probability distribution is known as the **hypergeometric distribution.**

It will be useful for our subsequent work on combining evidence from more than one 2×2 table in Chapter 13 to refer to the expected value and variance of the hypergeometric distribution. These are given as follows:

EQUATION 10.9

Expected Value and Variance of the Hypergeometric Distribution Suppose we consider all possible tables with fixed row margins N_1, N_2 and fixed column margins M_1, M_2, where $N_1 \le N_2$ and $M_1 \le M_2$ and $N = N_1 + N_2 = M_1 + M_2$. Let the random variable X denote the cell count in the (1, 1) cell. The expected value and variance of X are given by

$$E(X) = \frac{M_1 N_1}{N}$$

$$Var(X) = \frac{M_1 M_2 N_1 N_2}{N^2(N-1)}$$

The basic strategy in testing the hypothesis H_0: $p_1 = p_2$ versus H_1: $p_1 \neq p_2$ will be to enumerate all possible tables with the same margins as the observed table and compute the exact probability for each such table based on the hypergeometric distribution. A method for accomplishing this task is given as follows:

EQUATION 10.10

Enumeration of All Possible Tables with the Same Margins as the Observed Table

(1) Rearrange the rows and columns of the observed table so that the smaller row total is in the first row and the smaller column total is in the first column.

Suppose that after the rearrangement, the cells in the observed table are a, b, c, d, as shown in Table 10.10.

(2) Start with the table with 0 in the (1, 1) cell. The other cells in this table are then determined from the row and column margins. Indeed, to maintain the same row and column margins as the observed table, the (1, 2) element must be $a + b$, the (2, 1) cell must be $a + c$, and the (2, 2) element must be $(c + d) - (a + c) = d - a$.

(3) Construct the next table by increasing the (1, 1) cell by 1 (i.e., from 0 to 1), decreasing the (1, 2) and (2, 1) cells by 1, and increasing the (2, 2) cell by 1.

(4) Continue increasing and decreasing the cells by 1, as in step 3, until one of the cells is 0, at which point all possible tables with the given row and column margins have been enumerated. Each table in the sequence of tables is referred to by its (1, 1) element. Thus, the first table is the 0 table, the next table is the 1 table, and so on.

Example 10.19 **Cardiovascular Disease, Nutrition** Enumerate all possible tables with the same row and column margins as the observed data in Table 10.9.

Solution The observed table has $a = 2$, $b = 23$, $c = 5$, $d = 30$. The rows or columns do not need to be rearranged, because the first row total is smaller than the second row total, and the first column total is smaller than the second column total. Start with the 0 table, which has 0 in the (1, 1) cell, 25 in the (1, 2) cell, 7 in the (2, 1) cell, and 30 – 2, or 28, in the (2, 2) cell. The 1 table then has 1 in the (1, 1) cell, 25 – 1 = 24 in the (1, 2) cell, 7 – 1 = 6 in the (2, 1) cell, and 28 + 1 = 29 in the (2, 2) cell. Continue in this fashion until the 7 table is reached, which has 0 in the (2, 1) cell, at which point all possible tables with the given row and column margins have been enumerated. The collection of tables and their associated probabilities based on the hypergeometric distribution in Equation 10.8 are given in Table 10.12.

The question now is what should be done with these probabilities to evaluate the significance of the results? The answer depends on whether a one-sided or a two-sided alternative is being used. In general, the following procedure can be used:

EQUATION 10.11

Fisher's Exact Test: General Procedure and Computation of *p*-Value To test the hypothesis H_0: $p_1 = p_2$ versus H_1: $p_1 \neq p_2$, where the expected value of at least one cell is <5 when the data are analyzed in the form of a 2 × 2 contingency table, use the following procedure:

TABLE 10.12 **Enumeration of all possible tables with fixed margins and their associated probabilities based on the hypergeometric distribution for Example 10.19**

0	25
7	28

.017

1	24
6	29

.105

2	23
5	30

.252

3	22
4	31

.312

4	21
3	32

.214

5	20
2	33

.082

6	19
1	34

.016

7	18
0	35

.001

(1) Enumerate all possible tables with the same row and column margins as the observed table, as shown in Equation 10.10.

(2) Compute the exact probability of each table enumerated in step 1, using either the computer or the method in Equation 10.7.

(3) Suppose that the observed table is the a table and that the last table enumerated is the k table.

(a) To test the hypothesis H_0: $p_1 = p_2$ versus H_1: $p_1 \neq p_2$, the p-value $= 2 \times \min[Pr(0) + Pr(1) + \cdots + Pr(a), Pr(a) + Pr(a+1) + \cdots + Pr(k), .5]$.

(b) To test the hypothesis H_0: $p_1 = p_2$ versus H_1: $p_1 < p_2$, the p-value $= Pr(0) + Pr(1) + \cdots + Pr(a)$.

(c) To test the hypothesis H_0: $p_1 = p_2$ versus H_1: $p_1 > p_2$, the p-value $= Pr(a) + Pr(a+1) + \cdots + Pr(k)$.

For each of these three alternative hypotheses, the p-value can be interpreted as the probability of obtaining a table as extreme as or more extreme than the observed table.

Example 10.20 **Cardiovascular Disease, Nutrition** Evaluate the statistical significance of the data in Example 10.17.

Solution Suppose there is a two-sided alternative of the form H_0: $p_1 = p_2$ versus H_1: $p_1 \neq p_2$. Our table is the 2 table whose probability is .252 in Table 10.12. Thus, to compute the p-value, the smaller of the tail probabilities corresponding to the 2 table is computed and doubled. This strategy corresponds to the procedures for the various normal-theory tests studied in Chapters 7 and 8. First compute the left-hand tail area,

$$Pr(0) + Pr(1) + Pr(2) = .017 + .105 + .252 = .375$$

and the right-hand tail area,

$$Pr(2) + Pr(3) + \cdots + Pr(7) = .252 + .312 + .214 + .082 + .016 + .001 = .877$$

Then $p = 2 \times \min(.375, .877, .5) = 2(.374) = .749$

If a one-sided alternative of the form H_0: $p_1 = p_2$ versus H_1: $p_1 < p_2$ is used, then the p-value equals

$$Pr(0) + Pr(1) + Pr(2) = .017 + .105 + .252 = .375$$

Thus the two proportions in this example are *not* significantly different with either a one-sided or two-sided alternative, and we *cannot* say, on the basis of this limited amount of data, that there is a significant association between salt intake and cause of death.

In most instances, computer programs are used to implement Fisher's exact test using statistical packages such as SAS. There are other possible approaches to significance testing in the two-sided case. For example, the approach used by SAS is to compute

$$p\text{-value (two-tail)} = \sum_{\{i:\, Pr(i) \le Pr(a)\}} Pr(i)$$

In other words, the two-tail p-value using SAS is the sum of the probabilities of all tables whose probabilities are ≤ the probability of the observed table. Using this approach, the two-tail p-value would be

$$\begin{aligned} p\text{-value (two-tail)} &= Pr(0) + Pr(1) + Pr(2) + Pr(4) + Pr(5) + Pr(6) + Pr(7) \\ &= .017 + .105 + .252 + .214 + .082 + .016 + .001 = .688 \end{aligned}$$

In this section, we learned about Fisher's exact test, which is used for comparing binomial proportions from two independent samples in 2 × 2 tables with small expected counts (<5). This is the two-sample analogue to the exact one-sample binomial test given in Equation 7.36. If we refer to the flowchart at the end of this chapter (p. 412), we would answer yes to (1) Are samples independent? and no to (2) Are all expected values ≥5? This leads us to the box entitled "Use Fisher's exact test."

SECTION 10.4 Two-Sample Test for Binomial Proportions for Matched-Pair Data (McNemar's Test)

Example 10.21 **Cancer** Suppose we want to compare two different chemotherapy regimens for breast cancer after mastectomy. The two treatment groups should be as comparable as possible on other prognostic factors. To accomplish this goal, a matched study is set up such that a random member of each matched pair gets treatment A (chemotherapy) perioperatively (i.e., within 1 week after mastectomy) and for an additional 6 months, whereas the other member gets treatment B (chemotherapy only perioperatively). The patients are assigned to pairs matched on age (within 5 years) and clinical condition. The patients are followed for 5 years, with survival as the outcome variable. The data are displayed in a 2 × 2 table, as shown in Table 10.13. Notice the small difference in survival between the two treatment groups; the 5-year survival rate for treatment A = 526/621 = .847 and for treatment B = 515/621 = .829. Indeed, the Yates-corrected chi-square statistic as given in Equation 10.5 is 0.59 with 1 *df*, which is not significant. However, *the use of this test is valid only if the two samples are independent.* From the manner in which the samples were selected it is obvious that they are *not* independent, because members of each matched pair are similar in age and clinical condition. Thus, the Yates-corrected chi-square test *cannot* be used in this situation, since the *p*-value will not be correct. How then can the two treatments be compared using a hypothesis test?

TABLE 10.13 **A 2 × 2 contingency table comparing treatments A and B for breast cancer based on 1242 patients**

Treatment	Outcome: Survive for 5 years	Outcome: Die within 5 years	Total
A	526	95	621
B	515	106	621
Total	1041	201	1242

Suppose a different kind of 2 × 2 table is constructed to display these data. In Table 10.13 the *person* was the unit of analysis, and the sample size was 1242 people. In Table 10.14 the *matched pair* is the unit of analysis and *pairs* are classified according to whether or not the members of that pair survived for 5 years. Notice that Table 10.14 has 621 units rather than the 1242 units in Table 10.13. Furthermore, there are 90 pairs in which both patients died within 5 years, 510 pairs in which both patients survived for 5 years, 16 pairs in which the treatment A patient survived and the treatment B patient died, and 5 pairs in which the treatment B patient survived and the treatment A patient died. The dependence of the two samples can be illustrated by noting that the probability that the treatment B member of the pair survived given that the treatment A member of the pair survived = 510/526 = .970, while the probability that the treatment B member of the pair survived given that the treatment A member of the pair died = 5/95 = .053. If the samples were independent, then these two probabilities should be about the same. Thus, we conclude that the samples are highly dependent and that the chi-square test cannot be used.

TABLE 10.14 **A 2 × 2 contingency table with the matched pair as the sampling unit based on 621 matched pairs**

Outcome of treatment A patient	Outcome of treatment B patient: Survive for 5 years	Outcome of treatment B patient: Die within 5 years	Total
Survive for 5 years	510	16	526
Die within 5 years	5	90	95
Total	515	106	621

In Table 10.14, for 600 pairs (90 + 510), the outcomes of the two treatments are the same, whereas for 21 pairs (16 + 5), the outcomes of the two treatments are different. The following special names are given to each of these types of pairs:

DEFINITION 10.2 A **concordant pair** is a matched pair in which the outcome is the same for each member of the pair.

DEFINITION 10.3 A **discordant pair** is a matched pair in which the outcomes are different for the members of the pair.

Example 10.22 | There are 600 concordant pairs and 21 discordant pairs for the data in Table 10.14.

The concordant pairs provide no information about *differences between treatments* and will not be used in the assessment. Instead, we will focus on the discordant pairs, which can be divided into two types:

DEFINITION 10.4 A **type A discordant pair** is a discordant pair in which the treatment A member of the pair has the event and the treatment B member does not. Similarly, a **type B discordant pair** is a discordant pair in which the treatment B member of the pair has the event and the treatment A member does not.

Example 10.23 | If we define having an event as dying within 5 years, then there are 5 type A discordant pairs and 16 type B discordant pairs from the data in Table 10.14.

Let p = the probability that a discordant pair is of type A. If the treatments are equally effective, then about an equal number of type A and type B discordant pairs would be expected, and p should $= \frac{1}{2}$. If treatment A is more effective than treatment B, then less type A than type B discordant pairs would be expected, and p should be $< \frac{1}{2}$. Finally, if treatment B is more effective than treatment A, then more type A than type B discordant pairs would be expected, and p should be $> \frac{1}{2}$.

Thus, we wish to test the hypothesis H_0: $p = \frac{1}{2}$ versus H_1: $p \neq \frac{1}{2}$.

10.4.1 Normal-Theory Test

Suppose that of n_D discordant pairs, n_A are type A. Then under H_0, $E(n_A) = n_D/2$ and $Var(n_A) = n_D/4$, from the mean and variance of a binomial distribution, respectively. We will assume that the normal approximation to the binomial distribution holds, but will use a continuity correction for a better approximation. This approximation will be valid if $npq = n_D/4 \geq 5$ or $n_D \geq 20$. The following test procedure, referred to as McNemar's test, can then be used:

EQUATION 10.12

McNemar's Test for Correlated Proportions—Normal-Theory Test

(1) Form a 2×2 table of matched pairs, where the outcomes for the treatment A members of the matched pairs are listed along the rows and the outcomes for the treatment B members are listed along the columns.

(2) Count the total number of discordant pairs (n_D) and the number of type A discordant pairs (n_A).

(3) Compute the test statistic

$$X^2 = \left(\left|n_A - \frac{n_D}{2}\right| - \frac{1}{2}\right)^2 \bigg/ \left(\frac{n_D}{4}\right)$$

An equivalent version of the test statistic is also given by

$$X^2 = \left(|n_A - n_B| - 1\right)^2 \big/ (n_A + n_B)$$

(4) For a two-sided level α test, if

$$X^2 > \chi^2_{1,1-\alpha}$$

then reject H_0; if $X^2 \le \chi^2_{1,1-\alpha}$

then accept H_0.

(5) The exact p-value is given by $p\text{-value} = Pr(\chi_1^2 \ge X^2)$.

(6) Use this test only if $n_D \ge 20$.

The acceptance and rejection regions for this test are depicted in Figure 10.5. The computation of the p-value for McNemar's test is depicted in Figure 10.6.

This is a two-sided test despite the one-sided nature of the critical region in Figure 10.5. The rationale for this is that if either $p < \frac{1}{2}$ or $p > \frac{1}{2}$, $|n_A - n_D/2|$ will be

FIGURE 10.5 Acceptance and rejection regions for McNemar's test—normal-theory method

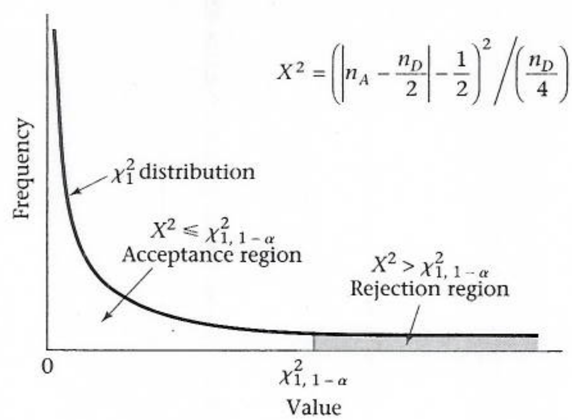
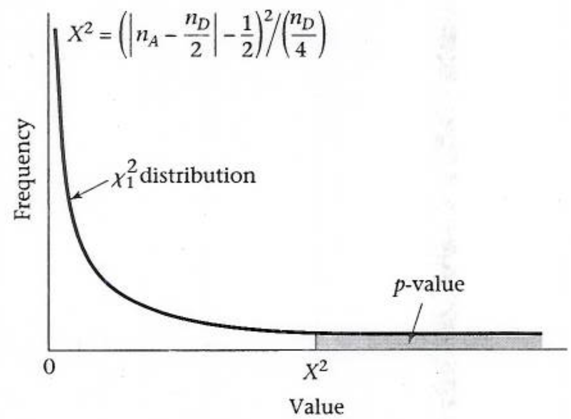


FIGURE 10.6 **Computation of the p-value for McNemar's test—normal-theory method**



large and, correspondingly, X^2 will be large. Thus, for alternatives on either side of the null hypothesis $\left(p = \frac{1}{2}\right)$, H_0 is rejected if X^2 is large and accepted if X^2 is small.

Example 10.24 **Cancer** Assess the statistical significance of the data in Table 10.14.

Solution Note that $n_D = 21$. Since $n_D\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 5.25 \ge 5$, the normal approximation to the binomial distribution is valid and the test in Equation 10.12 can be used. We have

$$X^2 = \frac{\left(|5 - 10.5| - \frac{1}{2}\right)^2}{21/4} = \frac{\left(5.5 - \frac{1}{2}\right)^2}{5.25} = \frac{5^2}{5.25} = \frac{25}{5.25} = 4.76$$

Equivalently, we could also compute the test statistic from

$$X^2 = \frac{\left(|5 - 16| - 1\right)^2}{5 + 16} = \frac{10^2}{21} = 4.76$$

From Table 6 in the Appendix, note that

$$\chi^2_{1,.95} = 3.84$$
$$\chi^2_{1,.975} = 5.02$$

Thus, because $3.84 < 4.76 < 5.02$, it follows that $.025 < p < .05$, and the results are statistically significant.

We conclude that *if the treatments give different results from each other* for the members of a matched pair, then the treatment A member of the pair is significantly more likely to survive for 5 years than the treatment B member. Thus, all other things being equal (such as toxicity, cost, etc.), treatment A would be the treatment of choice.

10.4.2 Exact Test

If $n_D/4 < 5$—that is, if $n_D < 20$—then the normal approximation to the binomial distribution cannot be used, and a test based on exact binomial probabilities is required. The details of the test procedure are similar to the one-sample binomial test in Equation 7.36 and are summarized as follows:

EQUATION 10.13

McNemar's Test for Correlated Proportions—Exact Test

(1) Follow the procedure in step 1 in Equation 10.12.

(2) Follow the procedure in step 2 in Equation 10.12.

(3) $p = 2 \times \sum_{k=0}^{n_A} \binom{n_D}{k} \left(\frac{1}{2}\right)^{n_D}$ if $n_A < n_D/2$

$p = 2 \times \sum_{k=n_A}^{n_D} \binom{n_D}{k} \left(\frac{1}{2}\right)^{n_D}$ if $n_A > n_D/2$

$p = 1$ if $n_A = n_D/2$

(4) This test is valid for any number of discordant pairs (n_D) but is particularly useful for $n_D < 20$, when the normal-theory test in Equation 10.12 cannot be used.

The computation of the p-value for this test is depicted in Figure 10.7.

Example 10.25 **Hypertension** A recent phenomenon in the recording of blood pressure is the development of the automated blood-pressure machine, where for a small fee a person can sit in a booth and have his or her blood pressure measured by a computer device. A study is conducted to compare the computer device with standard methods of measuring blood pressure. Twenty patients are recruited, and their hypertensive status is assessed by both the computer device and a trained observer. Hypertensive status is defined as either hypertensive (+), if either systolic blood pressure ≥ 160 or diastolic blood pressure ≥ 95, or normotensive (–) otherwise. The data are given in Table 10.15. Assess the statistical significance of these findings.

Solution An ordinary Yates-corrected chi-square test cannot be used for these data, because each person is being used as his or her own control and there are *not* two independent samples. Instead, a 2 × 2 table of matched pairs is formed, as shown in Table 10.16. Note that 3 people are measured as hypertensive by both the computer device and the trained observer, 9 people are normotensive by both methods, 7 people are hypertensive by the computer device and normotensive by the trained observer, and 1 person is normotensive by the computer device and hypertensive by the trained observer. Therefore, there are 12 (9 + 3) concordant pairs and 8 (7 + 1) discordant pairs (n_D). Because $n_D < 20$, the exact method must be used. We see that $n_A = 7$, $n_D = 8$. Therefore, because $n_A > n_D/2 = 4$, it follows from Equation 10.13 that

$$p = 2 \times \sum_{k=7}^{8} \binom{8}{k} \left(\frac{1}{2}\right)^{8}$$

FIGURE 10.7 Computation of the *p*-value for McNemar's test—exact method

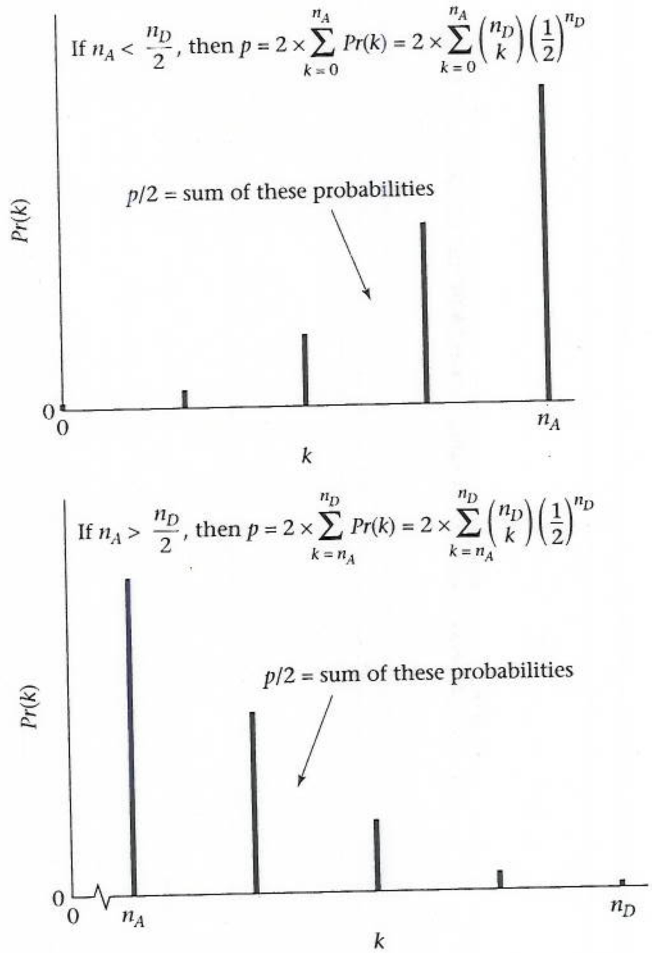


TABLE 10.15 Hypertensive status of 20 patients as judged by a computer device and a trained observer

	Hypertensive status			Hypertensive status	
Person	Computer device	Trained observer	Person	Computer device	Trained observer
1	−	−	11	+	−
2	−	−	12	+	−
3	+	−	13	−	−
4	+	+	14	+	−
5	−	−	15	−	+
6	+	−	16	+	−
7	−	−	17	+	−
8	+	+	18	−	−
9	+	+	19	−	−
10	−	−	20	−	−

TABLE 10.16 **Comparison of hypertensive status as judged by a computer device and a trained observer**

Computer device	Trained observer +	Trained observer −
+	3	7
−	1	9

This expression can be evaluated using Table 1 in the Appendix by referring to $n = 8$, $p = .5$ and noting that $Pr(X \geq 7|p = .5) = .0313 + .0039 = .0352$. Thus, the two-tailed p-value $= 2 \times .0352 = .070$.

Alternatively, a computer program could be used to perform the computations, as shown in Table 10.17. Note that the first and second columns have been interchanged so that the discordant pairs appear in the diagonal elements (and are easier to identify). In summary, the results are not statistically significant, and we cannot conclude that there is a significant difference between the two methods, although a *trend* toward the computer device identifying more hypertensives than the trained observer can be detected.

TABLE 10.17 **Use of SPSS^X/PC McNemar's test program to evaluate the significance of the data in Table 10.16**

```
                          SPSSX/PC  Release 1.0
-----McNemar Test
     COMP      COMPUTER DEVICE
with OBS       TRAINED OBSERVER

                          OBS
                     2.00      1.00          Cases             20
                  +---------+---------+
             1.00 |    7    |    3    |
COMP              +---------+---------+      (Binomial)
             2.00 |    9    |    1    |      2-tailed P     0.0703
                  +---------+---------+
```

Note that for a two-sided one-sample binomial test, the hypothesis H_0: $p = p_0$ versus H_1: $p \neq p_0$ is tested. In the special case where $p_0 = 1/2$, the same test procedure as for McNemar's test is also followed.

Finally, if we interchange the designation of which of 2 outcomes is an event, then the p-values will be the same in Equations 10.12 and 10.13. For example, if we define an event as surviving for 5+ years, rather than dying within 5 years in Table 10.13, then $n_A = 16$, $n_B = 5$ (rather than $n_A = 5$, $n_B = 16$ in Example 10.23). However, the test statistic X^2 and the p-value are the same because $|n_A - n_B|$ remains the same in Equation 10.12. Similarly, the p-value remains the same in Equation 10.13 due to the symmetry of the binomial distribution when $p = 1/2$ (under H_0).

In this section, we have studied McNemar's test for correlated proportions, which is used to compare two binomial proportions from matched samples. We studied both a large-sample test when the normal approximation to the binomial distribution is valid (i.e., when the number of discordant pairs, n_D, ≥ 20) and a small-sample test when $n_D < 20$. Referring to the flowchart at the end of this chapter (p. 412), we would answer no to (1) are samples independent? which would lead us to the box entitled "Use McNemar's test."

SECTION 10.5 Estimation of Sample Size and Power for Comparing Two Binomial Proportions

In Section 8.10, methods for estimating the sample size needed to compare means from two normally distributed populations were presented. In this section, similar methods for estimating the sample size required to compare two proportions are developed.

10.5.1 Independent Samples

Example 10.26 **Cancer, Nutrition** Suppose we know from Connecticut tumor-registry data that the incidence rate of breast cancer over a 1-year period for initially disease-free women ages 45–49 is 150 cases per 100,000 [2]. We wish to study whether or not the ingestion of large doses of vitamin A in tablet form will prevent breast cancer. The study is set up with (1) a control group of 45- to 49-year-old women who are given placebo pills by mail and are anticipated to have the same disease rate as indicated in the Connecticut tumor-registry data and (2) a study group of similarly aged women who are given vitamin A pills by mail and are anticipated to have a 20% reduction in risk. How large a sample is needed if a two-sided test with a significance level of .05 is used and a power of 80% is desired?

We wish to test the hypothesis H_0: $p_1 = p_2$ versus H_1: $p_1 \neq p_2$. Suppose that we wish to conduct a test with significance level α and power $1 - \beta$ and anticipate that there will be k times as many people in group 2 as in group 1; that is, $n_2 = kn_1$. The sample size required in each of the two groups to achieve these objectives is given as follows:

EQUATION 10.14

Sample Size Needed to Compare Two Binomial Proportions Using a Two-Sided Test with Significance Level α and Power $1 - \beta$, Where One Sample (n_2) Is k Times as Large as the Other Sample (n_1) (Independent-Sample Case) To test the hypothesis H_0: $p_1 = p_2$ versus H_1: $p_1 \neq p_2$ for the specific alternative $|p_1 - p_2| = \Delta$, with a significance level α and power $1 - \beta$, the following sample size is required

$$n_1 = \left[\sqrt{\bar{p}\bar{q}\left(1+\frac{1}{k}\right)}z_{1-\alpha/2} + \sqrt{p_1q_1 + \frac{p_2q_2}{k}}z_{1-\beta}\right]^2 \Big/ \Delta^2$$

$$n_2 = kn_1$$