Chapter 2

Hypothesis Testing

2.1 Study Suggestions

The analysis of Chapter 1 is called *descriptive*. Chapter 2 introduces **statistical inference**, in particular, hypothesis testing.

Chapter 2 is perhaps the most challenging chapter in the text. Chapter 2 contains a great many important ideas and a substantial amount of new terminology. I encourage my students by emphasizing that these important new ideas will appear repeatedly throughout the text. Thus, it is not essential that these ideas be mastered during this first exposure.

In Chapter 1 you learned how to perform a CRD. The end result of the study is a 2×2 table of observed frequencies. For example, the Infidelity study described in the text in Chapter 1 yielded the table below.

Version	S	F	Total
1	7	3	10
2	4	6	10
Total	11	9	20

Of Therese's friends who were told the husband was cheating (version 1), 70 percent said they would tell the wife if asked. Of Therese's friends who were told the wife was cheating, however, only 40 percent said they would tell the husband if asked. Given the results of Therese's study, there are two natural questions:

1. Is the difference between versions of 30 percentage points "real?"

2. If the difference of 30 percentage points is "real," is it important?

I put the word real in quotation marks because it is not obvious what I mean by that word. Hypothesis testing can be viewed as a technical device for obtaining a quantitative, objective answer to the first question for a particular definition of the term "real."

The second question above is subjective, and any serious attempt to answer it must necessarily depend on the respondent's expertise in the subject area of the study. The space I can devote in a general text to other subject areas is severely limited, so I cannot give the second question a thorough answer for every study in the text. The second question is, however, discussed briefly throughout the text, beginning in Chapter 7, under the heading *practical importance.*

The debate between the Skeptic and the Advocate, I believe, is a useful metaphor for what statisticians mean by saying a difference is real. More precisely, the Advocate says the difference is real and the Skeptic says it is not. Hypothesis testing quantifies the debate between the Skeptic and Advocate, and this quantification is called objective because it depends on the data collected and not on the opinion of the researcher or others. (In *The Mismeasure of Man* by Stephen Jay Gould, (Norton, 1981) the author convincingly argues how difficult it is to be objective, and how hard we must strive even to approach objectivity in our thinking. In addition, I understand that, from the point of view of a student, "quantifying a debate" is not necessarily a natural or good thing to do. I am reminded of the poetry text referenced in the movie *The Dead Poet's Society* that explained how to quantify the value of a poem! I happen to believe that there is great value to quantifying the debate between the Skeptic and Advocate. As you work through the text, I hope that you will be convinced of the validity of my viewpoint.)

Recall that the Skeptic's Argument is presented in the text. It is important to keep separate the following two notions. First, it is important to understand what the Skeptic's argument becomes when applied to a particular CRD. Second, it is important to decide whether you consider the Skeptic's Argument to be reasonable for a particular set of data.

For example, consider the Infidelity study. First, the Skeptic argues that there are no subjects for whom the treatment matters; there is one group of subjects who would tell on either a male or female, and there is another group of subjects who would not tell, regardless of the gender of the cheater. (If the treatment *does* matter for a particular subject, then that subject would give different responses when contemplating male and female cheaters.)

Second, you must decide whether you agree with the Skeptic for the Infidelity study. You are free to make this decision however you want; in Chapter 2 you will learn how a statistician quantifies the issue of whether to believe the Skeptic.

It is convenient to use the Skeptic's Argument to introduce hypothesis testing. Each hypothesis test has four steps. The first step is the specification of the two hypotheses—the null, H_0, and the alternative, H_1. In the current setting the null hypothesis is that the Skeptic is correct and the alternative specifies the way in which the Skeptic is incorrect.

It will be helpful to have the hypotheses defined in terms of symbols. In Statistics, frequently we will consider experiments that were never performed. Curiously, often this proves helpful. For example, consider a study in which every subject is assigned to the first treatment. For this study, let p_1 denote the proportion of subjects who would give a success. Note that we do not know the numerical value of p_1 nor do we have any hope of knowing it. Next, consider the study in which every subject is assigned to the second treatment. For this study, let p_2 denote the proportion of subjects who would give a success. Again, we do not know the numerical value of p_2 nor do we have any hope of knowing it. Note, however, that on the assumption the Skeptic is correct, $p_1 = p_2$ (make sure you understand why). Thus, we can write the null hypothesis as

$$H_0 : p_1 = p_2.$$

If the null hypothesis is false, then either $p_1 > p_2$ or $p_1 < p_2$. This gives rise to three possibilities for the alternative hypothesis H_1.

- $p_1 > p_2$. This is called the first alternative or the '$>$' alternative.
- $p_1 < p_2$. This is called the second alternative or the '$<$' alternative.
- $p_1 > p_2$ or $p_1 < p_2$. This alternative is usually written as $p_1 \neq p_2$. This is called the third alternative or the '$\neq$' alternative.

How does one select from these three choices? There is only one absolute rule.

- The alternative must be selected before any data are collected.

(The reason for this rule will become clear later.)

Without a strong reason to behave otherwise, I always choose the third alternative because it includes both directions in which the Skeptic can be mistaken. So, what constitutes a strong reason? The text presents two paradigms, "Inconceivable" and "What happens next?" I will discuss only the former here.

Consider the Crohn's study. The alternative $p_1 < p_2$ states that the first treatment, cyclosporine, is inferior to the second treatment, the placebo. In my experience, medical researchers consider such an alternative to be "inconceivable," and would, thus, exclude it from their alternative and select the first alternative.

Similarly, in the Ballerina study, Julie thought it was "inconceivable" for $p_1 > p_2$ (what does this mean in words?). Thus, Julie chose the second alternative, $p_1 < p_2$.

In the Infidelity study, Therese considered each possibility to be "conceivable," and, hence, she used the third alternative.

We don't get to "see" the actual values of p_1 or p_2; in other words, we will never know with certainty which hypothesis is true. What we do get to see are the data, which allow us to compute $\hat{p}_1$ and $\hat{p}_2$.

For the Infidelity study $\hat{p}_1$ is larger than $\hat{p}_2$, which provides evidence that the Skeptic is wrong. My language reflects the approach taken by statisticians. We begin by assuming that the Skeptic is correct. The data provide evidence that the Skeptic is incorrect, but we refer to it by talking about the evidence in support of the alternative. Remember, we begin by assuming the null hypothesis is true, and we want to determine how strongly the data support the alternative. In what follows, we will learn how statisticians measure the strength of that evidence.

Step 2 of a hypothesis test is called *the test statistic and its sampling distribution*. The test statistic half is easy; it is the number that summarizes the evidence in the data. For the current test, called Fisher's test, the observed value of the test statistic is

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

The observed value of the test statistic is also called the actual value of the test statistic. Why do we use these adjectives—observed and actual? We use them because, as you will see, we need to distinguish the actual value of the test statistic from lots of pretend values of the test statistic.

To summarize, for the Infidelity study, $x = 0.30$ reflects the evidence in support of the alternative; but how strong is this evidence?

I will digress by considering how I make assessments in everyday life. Many years ago, a physician told me the weight of my two-week-old son. I was woefully unprepared to interpret the number reported to me. Upon reflection, I realized my problem; for my entire life I had pretty much ignored the weights of two-week-old humans. In statistical jargon, I needed a reference group. The obvious reference group was all two-week-old babies; that is, I could hope to compare Roger's weight to the weights of all members of the reference group. The collection of all these weights (one per member of the reference group) is called the reference distribution. As I recall, I discovered that Roger's weight was "at the 95th percentile." This means that, roughly speaking, 95 percent of the numbers in the reference distribution for weight were less than or equal to Roger's weight. Alternatively, roughly speaking, 5 percent of the numbers in the reference distribution for weight were greater than or equal to Roger's weight. (We can be no more precise than saying 'roughly speaking' because some of the children had exactly the same weight as Roger. Percentiles arise in Chapter 12.)

A hypothesis test needs a reference group and reference distribution (the reference distribution is called the sampling distribution). In my life, my son was the special two-week-old and the collection of all two-week-olds formed the reference group. In the Infidelity study, we proceed as follows.

We view the data as arising from an assignment. When Therese performed her study, the process of randomization provided her with a particular assignment of subjects to treatments. But the actual assignment Therese obtained was just one of many possible assignments of subjects to treatments. Therese's actual assignment plays the role my son Roger played earlier, and the collection of all possible assignments that Therese could have obtained plays the role that all two-week-olds played earlier. In other words, the collection of all possible assignments is the reference group.

The reference, or sampling, distribution is simply the collection of the values of x that would have been obtained for each possible assignment. Here we run into a problem. Therese only did one experiment, she looked at only one assignment. Thus, the natural reaction is to note that we don't really know what data would have been obtained from other assignments. This impasse is overcome, however, by remembering that we are assuming that the null hypothesis is true, i.e. that the Skeptic is correct. Given that the Skeptic is correct, Therese had two groups of subjects: 11 subjects who would tell regardless of their treatment, and nine subjects who would not tell regardless of their treatment. Given this assumption, it is possible to determine the value of x for every possible assignment.

(Technical note: For the Infidelity study it can be shown that there are 184,756 different assignments of subjects to treatments; far too many to examine individually! Refer to the Class study in the text for a more managable example (with only six possible assignments). For more details, see the *More Mathematics* section of this chapter.)

The sampling distribution can be presented as a table with two columns. For example, Table 2.1 is the sampling distribution of the test statistic for the Infidelity study.

Even though there are 184,756 different assignments of the subjects to the treatments, there are only

Table 2.1: The sampling distribution of the test statistic for Fisher's test for the Infidelity study.

x	$P(X = x)$
-0.90	0.0001
-0.70	0.0027
-0.50	0.0322
-0.30	0.1500
-0.10	0.3151
0.10	0.3151
0.30	0.1500
0.50	0.0322
0.70	0.0027
0.90	0.0001
Total	1.0002

Table 2.2: The sampling distribution of the test statistic for Fisher's test for the Infidelity study.

x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.90	0.0001	0.0001	1.0000
-0.70	0.0027	0.0027	0.9999
-0.50	0.0322	0.0349	0.9973
-0.30	0.1500	0.1849	0.9651
-0.10	0.3151	0.5000	0.8151
0.10	0.3151	0.8151	0.5000
0.30	0.1500	0.9651	0.1849
0.50	0.0322	0.9973	0.0349
0.70	0.0027	0.9999	0.0027
0.90	0.0001	1.0000	0.0001

ten different values of x, and they are listed in the first column of the table. Note that, by convention, we list the possible values of x in order from smallest to largest. (The farther a number is to the left on the number line, the smaller it is.) Notice that the actual value of the test statistic, 0.30, is in this list. The numbers in the second column equal the proportion of assignments that would give the corresponding value of the test statistic. Note that these proportions are rounded, which is why the total is 1.0002 instead of 1.0000. For example, it can be shown that 11 assignments yield $x = -0.90$;, thus, the first entry in the second column equals 0.0001, the number one gets upon rounding $11/184{,}756 = 0.000060$, to the nearest ten-thousandth.

On occasion, a statistician will need the cumulative sums of the second column (the proportions) from top-to-bottom and from bottom-to-top. Thus, typically we present sampling distributions with these two additional columns. For example, for the Infidelity study the sampling distribution typically is presented as in Table 2.2. The third column is the cumulative sum of the entries in the second column, working down the column. The fourth column is the cumulative sum of the entries in the second column, working up the column. For example, the second entry in the third column, 0.0027, equals the proportion of all assignments that yield -0.70 or smaller for the value of x. This means $x = -0.70$ or $x = -0.90$. This proportion is obtained by adding the proportion that gives $x = -0.90$ to the proportion that gives $x = -0.70$; in other words, the first two entries in column 2. Thus, the 0.0027 in row 2, column 3 equals the sum of the first two entries in column 2, namely $0.0001 + 0.0027$. The minor lack of agreement, the sum is 0.0028 and the entry is 0.0027, is caused by round-off error. The entry is the more accurate value. (Note: As stated earlier the first entry in column 2 is actually $11/184{,}756$. The second entry in that column is actually $495/184{,}756$. Summing these gives $506/184{,}756{=}0.002739$, which is rounded to give the tabled value, 0.0027.)

The text states that the remaining two steps of hypothesis testing are the rule of evidence and the P-value. It is my experience that it is easier to combine these steps.

Therese's alternative hypothesis states $p_1 \neq p_2$. It is useful to break this into two disjoint pieces, $p_1 > p_2$ and $p_1 < p_2$. In the Infidelity study, $\hat{p}_1 = 0.70$ is larger than $\hat{p}_2 = 0.40$; thus, intuitively, of the two pieces of the alternative, the data support $p_1 > p_2$ more than they support $p_1 < p_2$. As a result, we are going to begin by focusing on the piece $p_1 > p_2$.

Consider the following question. Which possible values of x support $p_1 > p_2$ more strongly than the actual value of $x = 0.30$ supports $p_1 > p_2$? Upon reflection, the answer is any value of x that is larger than 0.30.

Now consider the other piece of the alternative,

$p_1 < p_2$. Upon reflection, you will note that the value $x = -0.30$ supports this piece of the alternative with exactly the same strength as 0.30 supports the other piece of the alternative, $p_1 > p_2$. Next, which possible values of x support $p_1 < p_2$ more strongly than $x = -0.30$ supports $p_1 < p_2$? Upon reflection, the answer is any value of x that is smaller than -0.30.

The P-value is equal to

$$P(X \geq 0.30) + P(X \leq -0.30) =$$

$$0.1849 + 0.1849 = 0.3698.$$

It will take some time to understand this, so let's focus on some of its features.

- The P-value is obtained by adding together two proportions, with one proportion corresponding to each piece of the alternative.

- The P-value equals the proportion of assignments with the following property:

  > The value of x for the assignment gives either the same evidence as the actual x or stronger evidence than the actual x.

  Recall that when we say evidence it is implicit that we mean evidence in support of the alternative.

For a second example, consider the Chronic Crohn's disease study. Table 2.3 presents the sampling distribution of the test statistic for this study. Recall that the actual value of x was 0.27 and that we are using the first alternative, $p_1 > p_2$. Repeating our logic from the previous example, we note that any value of x larger than 0.27 would provide even stronger evidence than the actual $x = 0.27$. Thus, the P-value equals

$$P(X \geq 0.27) = 0.0198.$$

Notice that there is only one term in the formula for the P-value because our alternative excludes one possibility (that $p_1 < p_2$).

As a final example, consider the Ballerina study. Table 2.4 presents the sampling distribution of the test statistic for this study. Recall that the actual

Table 2.3: The sampling distribution of the test statistic for Fisher's test for the Chronic Crohn's Disease study.

x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.46	0.0001	0.0001	1.0000
-0.41	0.0005	0.0006	0.9999
-0.35	0.0025	0.0031	0.9994
-0.29	0.0092	0.0123	0.9969
-0.24	0.0265	0.0388	0.9877
-0.18	0.0605	0.0993	0.9612
-0.12	0.1102	0.2095	0.9007
-0.07	0.1605	0.3700	0.7905
-0.01	0.1872	0.5572	0.6300
0.05	0.1752	0.7323	0.4428
0.10	0.1314	0.8637	0.2677
0.16	0.0788	0.9425	0.1363
0.21	0.0377	0.9802	0.0575
0.27	0.0143	0.9945	0.0198
0.33	0.0043	0.9988	0.0055
0.38	0.0010	0.9998	0.0012
0.44	0.0002	1.0000	0.0002

Table 2.4: The sampling distribution of the test statistic for Fisher's test for the Ballerina study.

x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.40	0.0009	0.0010	1.0000
-0.32	0.0081	0.0090	0.9990
-0.24	0.0387	0.0477	0.9910
-0.16	0.1127	0.1604	0.9523
-0.08	0.2104	0.3708	0.8396
0.00	0.2584	0.6292	0.6292
0.08	0.2104	0.8396	0.3708
0.16	0.1127	0.9523	0.1604
0.24	0.0387	0.9910	0.0477
0.32	0.0081	0.9990	0.0090
0.40	0.0009	1.0000	0.0010

value of x was -0.24 and that we are using the second alternative, $p_1 < p_2$. Repeating our logic from the previous examples, we note that any value of x smaller than -0.24 would provide even stronger evidence than the actual $x = -0.24$. Thus, the P-value equals

$$P(X \leq -0.24) = 0.0477.$$

Notice that there is only one term in the formula for the P-value because our alternative excludes one possibility (that $p_1 > p_2$).

Before a chance mechanism is operated, often we want to be able to say something about the outcome that will result. The language of this “saying something” is probability. Remember that probabilities are computed before the chance mechanism is allowed to operate.

It is important, for your general understanding as well as for the goals of the text, that you understand the long-run relative frequency interpretation of probability. In particular, this interpretation is central to the development of simulation approximations that are introduced in Chapter 3. In order to understand the long-run relative frequency interpretation, you must understand the two kinds of chance mechanisms, namely those that can be operated repeatedly under identical conditions and those that can not.

The rules of probability do not play an important role in this text.

The most time-consuming part of performing Fisher’s test is the determination of the sampling distribution of the test statistic. Do not despair; a simple method for obtaining an approximate answer is given in Chapter 3. Until then, remember that the sampling distribution is always given to you (unless you are asked to obtain it by brute force for a study with a very small number of subjects).

Many of my students have difficulty with Exercise 9 on page 74 of the text. This question asks for the P-value for the third alternative for a particular unbalanced CRD. If the test statistic is $x = 0.35$ then, literally, the P-value is

$$P(X \geq 0.35) + P(X \leq -0.35).$$

The first term in the above sum is no problem; its value, 0.0712, can be read directly from Table 2.16 on page 74, which presents the sampling distribution of X. The second term in the sum, however, is troublesome; there is no row in the table corresponding to $x = -0.35$ because -0.35 is not a possible value of the test statistic. This difficulty provides a good example of the importance of thinking about what a formula *means* rather than mechanically following an algorithm that yields an answer. Because the test statistic has for its possible values

$$-0.55, -0.40, -0.25, \ldots, \text{ and } 0.95,$$

the event $(X \leq -0.35)$ is identical to the event $(X \leq -0.40)$. The probability of this latter event can be obtained from Table 2.16; it is 0.0369.

The remainder of this section discusses two subtle points that you can ignore if that is your preference.

- On page 57 of the text, I argue that peeking, as I call it, to determine the values of m_1 and m_2 is allowable. My favorite argument is the one given in the book, namely that this peeking gives no insight into whether the performances of the two treatments were substantially the same or grossly different. Occasionally, this argument does not satisfy a student. In such situations, I argue as follows.

  Consider the class study. Before collecting data, the researcher knows that there are five possible values for m_1, namely 0, 1, 2, 3, and 4. The researcher could obtain the sampling distribution of X for each of these five possibilities. Once the data are collected, however, the researcher would realize that all but one of these sampling distributions can be disregarded. For example, if the actual study gives $m_1 = 2$, then according to the null hypothesis, two subjects are type A and two are type D. Thus, every possible assignment of subjects to treatments would yield $m_1 = 2$. Peeking is simply a labor-saving device—it saves us from wasting time computing sampling distributions that we will not need.

- There seems to be an inconsistency in my arguments. On the one hand, I emphasize that probabilities must be computed *before* the operation of the chance mechanism. On the other hand, the formula for computing the P-value involves a number, x, whose value is not known until *after* the chance mechanism of randomization has operated. This apparent inconsistency turns out not to be a problem, that is, it does not yield a misleading result, because we decide on the *form* of the event whose probability will be computed *before* the chance mechanism operates. For example, for the first alternative, we decide to compute $P(X \geq x)$ before collecting any data.

2.2 Solutions to Odd-Numbered Exercises

Solutions for Sections 2.1 and 2.2

3. A type A subject would yield a success on either treatment. Since the study yielded no successes, n_A must be 0.

5. A type C subject would yield a failure on treatment 1 or a success on treatment 2. Since treatment 1 yielded no failures, and treatment 2 no successes, n_C must be 0.

Solutions for Section 2.5

1. (a) The stronger evidence is provided by 0.50 because it is larger than 0.30.
   (b) The stronger evidence is provided by 0.30 because it is smaller than 0.50.
   (c) The stronger evidence is provided by 0.50 because it is farther from 0 than is 0.30.

3. (a) The stronger evidence is provided by -0.30 because it is larger than -0.50.
   (b) The stronger evidence is provided by -0.50 because it is smaller than -0.30.
   (c) The stronger evidence is provided by -0.50 because it is farther from 0 than is -0.30.

5. (a) For the first alternative,
   P $= P(X \geq 0.24) = 0.0782$;
   P $= P(X \geq 0.08) = 0.3888$.
   (b) For the second alternative,
   P $= P(X \leq -0.32) = 0.0232$;
   P $= P(X \leq 0.08) = 0.8019$.
   (c) For the third alternative,
   P $= P(X \geq 0.48) + P(X \leq -0.48) = 0.0008 + 0.0008 = 0.0016$;
   P $= P(X \geq 0.08) + P(X \leq -0.08) = 0.3888 + 0.3888 = 0.7776$.

9. (a) For the first alternative,
   P $= P(X \geq 0.35) = 0.0712$;
   P $= P(X \geq 0.50) = 0.0116$.
   (b) For the second alternative,
   P $= P(X \leq -0.25) = 0.1753$;
   P $= P(X \leq -0.40) = 0.0369$.
   (c) For the third alternative,
   P $= P(X \geq 0.65) + P(X \leq -0.65) = 0.0010 + 0.0000 = 0.0010$;
   P $= P(X \geq 0.40) + P(X \leq -0.40) = 0.0116 + 0.0369 = 0.0485$.

Solutions for Section 2.6

1. For the first alternative, $R = 10$, $C = 8$, and $O = 7$. Thus, **P** $= 0.0099$.

   For the third alternative, $R = 10$, $C = 8$, and $O = 7$. Thus, **P** $= 2(0.0099) = 0.0198$.

3. For the second alternative, $R = 11$, $C = 9$, and $O = 6$. Thus, **P** $= 0.1935$.

   For the third alternative, $R = 11$, $C = 9$, and $O = 6$. Thus, **P** $= 2(0.1935) = 0.3870$.

5. For the first alternative, $R = 5$, $C = 5$, and $O = 4$. Thus, **P** $= 0.1032$.

2.3 Exam Questions

1. Recall that the Skeptic's and Advocate's arguments were introduced on page 39 of the text. In the Infidelity study, a subject named Julie read the first version and responded yes.

   (a) True or false? According to the Skeptic's Argument, if Julie had read the second version then she would have responded no.

(b) True or false? According to the Advocate's Argument, if Julie had read the second version then she would have responded no.

2. A balanced CRD on 14 subjects yields $a = 1$ and $m_1 = 6$. Find the exact P-value for the second alternative (<).

3. The sampling distribution of the test statistic for Fisher's test for an unbalanced CRD is given by the following table.

x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.7	0.0186	0.0186	1.0000
-0.4	0.1632	0.1818	0.9814
-0.1	0.3916	0.5734	0.8182
0.2	0.3263	0.8998	0.4266
0.5	0.0932	0.9930	0.1002
0.8	0.0070	1.0000	0.0070

(a) What is the probability that the test statistic achieves the value -0.4?

(b) Find the P-value for the first alternative if $x = 0.5$.

(c) Find the P-value for the third alternative if $x = 0.5$.

4. Marge performs a CRD with 20 subjects on the first treatment, 10 subjects on the second treatment, and obtains a total of 16 successes. Marge determines the sampling distribution of the test statistic for Fisher's test, and it is given below.

x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.70	0.0003	0.0003	1.0000
-0.55	0.0053	0.0056	0.9997
-0.40	0.0390	0.0446	0.9944
-0.25	0.1386	0.1832	0.9554
-0.10	0.2668	0.4500	0.8168
0.05	0.2911	0.7410	0.5500
0.20	0.1819	0.9229	0.2590
0.35	0.0640	0.9869	0.0771
0.50	0.0120	0.9989	0.0131
0.65	0.0011	1.0000	0.0011
0.80	0.0000	1.0000	0.0000

(a) What would be the observed value of the test statistic if all 16 successes are obtained by subjects that have been assigned to the first treatment?

(b) Note that according to the sampling distribution, $P(X = 0.80) = 0.0000$.

True or false? This means that, according to the null hypothesis, no assignments would yield 0.80 as the observed value of the test statistic

(c) Find the P-value for the first alternative (>) if the test statistic equals 0.50.

(d) Find the P-value for the third alternative (≠) if the test statistic equals -0.55.

5. A CRD yields the following 2×2 table.

Treatment	S	F	Total
1	2	12	14
2	8	6	14
Total	10	18	28

Find the exact P-value for the third alternative (≠).

6. A balanced CRD has eight subjects, four men and four women. You may use the fact that there are 70 possible assignments of subjects to treatments.

(a) Compute the probability that the four women are assigned to the same treatment.

(b) The researcher performs the CRD, analyzes the data with Fisher's test, and obtains an exact P-value of 0.001. What should the researcher conclude?

7. Homer performs a CRD with 20 subjects on the first treatment, 25 subjects on the second treatment, and obtains a total of 12 successes. Homer determines the sampling distribution of the test statistic for Fisher's test, and it is given below.

x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.48	0.0002	0.0002	1.0000
-0.39	0.0031	0.0033	0.9998
-0.30	0.0216	0.0249	0.9967
-0.21	0.0810	0.1059	0.9751
-0.12	0.1822	0.2881	0.8941
-0.03	0.2591	0.5472	0.7119
0.06	0.2387	0.7859	0.4528
0.15	0.1432	0.9291	0.2141
0.24	0.0554	0.9845	0.0709
0.33	0.0134	0.9979	0.0155
0.42	0.0019	0.9999	0.0021
0.51	0.0001	1.0000	0.0001
0.60	0.0000	1.0000	0.0000

(a) What would be the observed value of the test statistic if a total of three successes is obtained by subjects that are assigned to the first treatment?

(b) There are approximately 3,170 billion possible assignments of subjects to treatments for this study. On the assumption that the null hypothesis is true, how many assignments would yield 0.06 as the value of the test statistic? (Choose the best answer from the choices below.)

i. 2,387
ii. 0.2387
iii. 757 billion
iv. 23.87 billion

8. A balanced CRD has four subjects, Alan, Brian, Carol, and Diana. Alan and Carol are husband and wife, and Diana and Brian are wife and husband. Let B denote the event that every subject is assigned to the same treatment as her/his spouse, and let C denote the event that every subject is assigned to the same treatment as the other subject of his/her sex.

   Compute $P(B \text{ or } C)$.

9. Peg performs a CRD with 40 subjects on the first treatment, and 50 subjects on the second treatment. She obtains a total of 10 successes. Peg determines the sampling distribution of the test statistic for Fisher's test, and it is given below.

x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
-0.200	0.0018	0.0018	1.0000
-0.155	0.0175	0.0193	0.9982
-0.110	0.0732	0.0925	0.9807
-0.065	0.1725	0.2650	0.9075
-0.020	0.2539	0.5189	0.7350
0.025	0.2437	0.7626	0.4811
0.070	0.1545	0.9171	0.2374
0.115	0.0639	0.9810	0.0829
0.160	0.0165	0.9975	0.0190
0.205	0.0024	0.9998	0.0025
0.250	0.0001	1.0000	0.0001

(a) Find the P-value for the first alternative ($>$) if the test statistic equals 0.160.

(b) Find the P-value for the third alternative ($\neq$) if the test statistic equals -0.110.

(c) Assuming the null hypothesis is true, is the test statistic more likely to positive, more likely to be negative, or equally likely to be positive or negative? Explain your answer.

(d) What would be the observed value of the test statistic if the treatments had an equal number of successes?

10. A chance mechanism has seven possible outcomes, denoted by

    b, c, d, e, f, g, and h.

    Assume that the equally likely case applies; that is, that the seven outcomes are equally likely to occur. For this chance mechanism:

    (a) Write down any event A with the property that the probability that A occurs equals $3/7$.

    (b) Write down any event B with the property that the probability that B occurs equals $1/2$.

11. A researcher conducts a balanced CRD with 12 subjects. You may use the fact that there are 924 different possible assignments of subjects to treatments.

    In addition, the study yields a total of five successes and seven failures.

(a) Given the above information, what is the largest possible value of the test statistic?

(b) How many different assignments give this largest possible value of the test statistic?

12. Recall that the Skeptic's Argument was introduced in Section 2.1 of the text. In the Three-Point Basket study, Clyde's first shot was a success from the left corner.

    True or false? According to the Skeptic's Argument, if Clyde's first shot had been from the front, then it would have been a success.

13. Recall that the Advocate's Argument was introduced in Section 2.1 of the text. In the Three-Point Basket study, Clyde's first shot was a success from the left corner.

    True or false? According to the Advocate's Argument, if Clyde's first shot had been from the front, then it could have been either a success or a failure.

14. Recall the following definition of subject type defined in Section 2.1:

	Response to	
Type	Treatment 1	Treatment 2
A	S	S
B	S	F
C	F	S
D	F	F

    True or false? If a subject yields a success on treatment 1, then the subject must be Type A or B.

15. Alice and Barbara are subjects in a CRD with two treatments and a dichotomous response. Both Alice and Barbara are assigned to the first treatment. Alice gives a response of success.

    True or false? If the null hypothesis is true and if she had been assigned to the second treatment, then Alice would have yielded a response of success.

16. Alice and Barbara are subjects in a CRD with two treatments and a dichotomous response. Both Alice and Barbara are assigned to the first treatment. Alice gives a response of success.

    True or false? If the null hypothesis is true, then Barbara must yield a response of success.

17. A researcher believes that it is "inconceivable" that treatment 1 is worse than treatment 2. What should the alternative hypothesis be?

18. Consider a balanced CRD with two treatments and a dichotomous response on subjects denoted by A, B, C, D, E, and F.

    (a) There are 20 possible assignments of subjects to treatments. List the 12 assignments that have the property that subjects A and B are assigned to different treatments.

    (b) Instead of performing a CRD, a researcher decides to select at random an assignment from one of the 12 you listed in (a). What is the probability that this researcher assigns subjects B and C to different treatments?

    (c) Refer to (b). The researcher conducts the study; the result is that subjects A, B, and C give successes while subjects D, E, and F give failures. Obtain the sampling distribution of $\hat{p}_1 - \hat{p}_2$.

19. Norm plans a balanced study on six subjects to compare two treatments with a dichotomous response. Denote the subjects by A, B, C, D, E, and F. Norm decides not to perform a CRD. In particular, he does not assign subjects to treatments by randomization. Instead, Norm has a box with 10 cards in it. Two cards in his box have ACE written on them, two cards have BDF written on them, and the six sequences of letters below appear on exactly one of Norm's cards.

    ACF, ADE, ADF, BCE, BCF, and BDE.

    Norm selects a card at random from his box. The subjects corresponding to the letters on the selected card are assigned to treatment 1, and the remaining subjects are assigned to treatment 2.

    Compute the probability that subjects A and C are assigned to the same treatment.

20. A political analyst states, "The probability is 0.4 that in the year 2012 the people of the United States will elect a woman to be president."

    True or false? The value of the probability, 0.4, in this statement can be interpreted by using the long-run relative frequency interpretation of probability.

21. A book on gambling states, "The probability that a craps player throws a total of nine with a pair of dice is $1/9$."

    True or false? The value of the probability, $1/9$, in this statement can be interpreted by using the long-run relative frequency interpretation of probability.

22. Bob states, "Every day I give my dog Casey either 1, 2, or 3 dog biscuits. The probability that I give her exactly one biscuit is 0.3, and the probability that I give her exactly two biscuits is 0.5." What is the probability that Bob gives Casey exactly three biscuits?

23. Bob states, "Every day I give my cat Roscoe either 1, 2, or 3 cat treats. The probability that I give him exactly one treat is 0.25, and the probability that I give him exactly two treats is 0.35." What is the probability that Bob gives Roscoe either one or two treats?

24. Bob and Roger are going to play three games of basketball. Each game will result in either Bob or Roger winning. Bob states, "The probability that I win all three games is 0.2, the probability that I win two of three games is 0.4, and the probability that I win a majority of the games is 0.7." Comment.

25. Sam cannot remember which of two numbers, 0.25 and 0.03, is his P-value, but he does remember that his data are not statistically significant. What is Sam's P-value?

26. You are given the following facts about a hypothesis test. (Note that this test is not a Fisher's test.)

    - The possible values of the test statistic W are 1, 2, . . . , 100.
    - If the null hypothesis is true, then the 100 possible values of W are equally likely.
    - The smaller the value of W, the stronger the evidence is in support of the alternative hypothesis.

    Given that the data yields $W = 7$, compute the P-value.

27. Consider a balanced study with eight subjects, identified as A, B, C, D, E, F. G, and H. In the actual study,

    - A, B, C, and D are assigned to the first treatment, and
    - There are exactly three successes, and they are obtained by A, B, and E.

    We have obtained the sampling distribution of the test statistic on the assumption that the Skeptic is correct. It also is possible to obtain a sampling distribution of the test statistic if the Skeptic is wrong *provided* we specify *exactly* how the Skeptic is in error. These new sampling distributions are used in the study of **statistical power** which is briefly described in Chapter 7 of the text.

    Assume that the Skeptic is correct about subjects A, D, E, and H, but is incorrect about the other four subjects (that is, for A, D, E, and H, each treatment would give the same response, but for the remaining subjects the different treatments would yield different responses). For the assignment that puts B, D, E, and F on the first treatment, determine the response for each of the eight subjects.

2.4 Solutions to Exam Questions

1. (a) False. (According to the Skeptic's Argument, Julie would have responded yes.)

   (b) False. (The Advocate's Argument does not specify how any particular individual will respond to the treatment not given.)

2. $R = 7$, $C = 6$, and $O = 5$. Thus, the P-value equals 0.0513.

3. (a) 0.1632.
   (b) $P(X \geq 0.5) = 0.1002$.
   (c) $P(X \geq 0.5) + P(X \leq -0.5) = 0.1002 + 0.0186 = 0.1188$.

4. (a) The 2×2 table is below.

Treatment	S	F	Total
1	16	4	20
2	0	10	10
Total	16	14	30

   Thus, $x = 0.80 - 0.00 = 0.80$.
   (b) False. (Part (a) shows that *some* assignments will yield $x = 0.80$; the proportion of such assignments is very small, however. In fact, to the nearest 0.0001 it is 0.0000.)
   (c) **P** $= P(X \geq 0.50) = 0.0131$.
   (d) **P** $= P(X \leq -0.55) + P(X \geq 0.65) = 0.0056 + 0.0011 = 0.0067$.

5. $R = 14$, $C = 10$, and $O = 8$, yielding Prob. $= 0.0230$. Thus, for the third alternative **P** $= 0.0460$.

6. (a) The four women could be assigned to the first treatment or to the second treatment; thus, the probability they are all assigned to the same treatment equals $2/70 = 0.0286$.
   (b) With 70 possible assignments, the smallest nonzero probability is $1/70$. The researcher should conclude that an error has been made; the exact P-value cannot be 0.001.

7. (a) The contingency table is below.

Treat.	S	F	Total
1	3	17	20
2	9	16	25
Total	12	33	45

$$x = \hat{p}_1 - \hat{p}_2 = 0.15 - 0.36 = -0.21.$$

   (b) The sampling distribution states that 23.87 percent of the assignments would yield 0.06 as the value of the test statistic, and 23.87 percent of 3,170 billion equals 757 billion.

8. As in the Class study, there are six possible equally likely assignments of subjects to treatments. Two assignments satisfy event B, namely the one that puts Alan and Carol to treatment 1, and the one that puts them on treatment 2. Similarly, the two assignments that put Alan and Brian together (on treatment 1 or 2) define event C. Thus,

$$P(B \text{ or } C) = \frac{2}{6} + \frac{2}{6} = \frac{4}{6}.$$

9. (a) The P-value equals

$$P(X \geq 0.160) = 0.0190.$$

   (b) The P-value equals

$$P(X \geq 0.110) + P(X \leq -0.110) =$$
$$P(X \geq 0.115) + P(X \leq -0.110) =$$
$$0.0829 + 0.0925 = 0.1754.$$

   (c) $P(X > 0) = P(X \geq 0.025) = 0.4811$; similarly, $P(X < 0) = 0.5189$. Thus, X is more likely to be negative than positive.
   (d) If each treatment had five successes, then $\hat{p}_1 = 5/40 = 0.125$, $\hat{p}_2 = 5/50 = 0.100$, and

$$x = 0.125 - 0.100 = 0.025.$$

10. (a) $A = \{$ b, c, d $\}$. Other answers are possible.
    (b) Impossible. The probability of B must be $0, 1/7, 2/7, 3/7, 4/7, 5/7, 6/7$, or 1.

11. (a) The largest possible value of the test statistic occurs when $a = 5$ and $c = 0$, These values give $\hat{p}_1 = 5/6$, $\hat{p}_2 = 0$, and $x = 5/6$.
    (b) Seven different assignments give this largest possible value of the test statistic because there are seven ways to choose the one failure that is assigned to the first treatment.

12. True.

13. True.

14. True.

15. True.

16. False.

17. The first alternative (>).

18. (a) To save space, an assignment is presented by listing the subjects it assigns to treatment 1. The 12 possible assignments are ACD, ACE, ACF, ADE, ADF, AEF, BCD, BCE, BCF, BDE, BDF, and BEF.
    (b) $6/12 = 0.5$.
    (c) The possible values of $\hat{p}_1 - \hat{p}_2$ are $1/3$ and $-1/3$; each of these values has probability equal to 0.5.

19. 0.6.

20. False.

21. True.

22. 0.2.

23. 0.60.

24. Bob's probabilities violate the third rule of probability. More precisely, the probability that Bob wins a majority of the games must equal the probability that he wins exactly two games (0.2) added to the probability that he wins three games (0.4).

25. 0.25.

26. 0.07.

27. This one is tricky. The answer is that D, G, and H yield failures and the remaining five subjects yield successes. The reasoning follows. B and D are still on the first treatment; thus, B will remain a success and D a failure. G and H are still on the second treatment; thus, they will both remain failures. A and C are switched from the first to the second treatment; A will remain a success because the Skeptic is correct, but C will switch to a success because the Skeptic is incorrect. Finally, E and F are switched from the second to the first treatment; E will remain a success because the Skeptic is correct, but F will switch to a success because the Skeptic is incorrect.

2.5 More Mathematics

In this section I will present a brief introduction to and derivation of the hypergeometric distribution.

On page 57 of the text, I state that there are 184,756 possible assignments of subjects to treatments for the Infidelity study. How do I know this? And how can mathematical statisticians study such a large number of assignments?

The starting point is a subject referred to as *Sophisticated Counting*. The first idea is the multiplication rule. Suppose your favorite fast food restaurant serves three kinds of sandwiches—hamburger, chicken, and soyburger. In addition, the restaurant has two choices of soft drink—Pepsi and Mountain Dew. Suppose you decide to order a meal consisting of one sandwich and one soft drink; how many different meals are possible? It is instructive to answer this question by drawing a diagram.

	Drink	
Sandwich	Pepsi	Mt. Dew
Hamburger		
Chicken		
Soyburger		

Note that there is a one-to-one correspondence between the empty cells in the above table and the possible meals. Simple counting reveals that there are six empty cells, and, hence, six possible meals. Because the empty cells form a rectangle, the answer, six, also can be obtained by multiplying 3 by 2.

The above fast food dilemma illustrates the multiplication rule. Suppose a complex choice consists of two component choices. Suppose further that the first component choice offers k_1 options, and that, regardless of the option selected by the first component choice, the second component choice offers k_2 options. Then the complex choice offers k_1k_2 options. For example, the complex choice of selecting a meal has two component choices—the choice of sandwich and the choice of drink. The first component choice has $k_1 = 3$ options, and regardless of the choice of the sandwich, the second component choice has $k_2 = 2$ options. Thus, there are $k_1 k_2 = 3(2) = 6$ possible meals.

The interested reader can prove the multiplication rule by mimicing the fast food problem by drawing a table with k_1 rows and k_2 columns.

The multiplication rule can be extended to any number of component choices. In order for the multiplication rule to be valid, however, remember that it is critical that the component choices can be set up so that the *number* of options available at any component choice is not affected by the particular option(s) selected by any earlier component choice(s). Some examples may help to clarify these ideas.

Suppose you decide to supplement your fast food meal with potatoes and a dessert. Suppose that your restaurant prepares potatoes in four different ways and offers a choice of five desserts. The multiplication rule applies and the number of possible meals is

$$3(2)(4)(5) = 120.$$

Suppose that you must select an outfit for a summer day. Your wardrobe consists of two pairs of blue jeans, two pairs of shorts, six tank tops, and six t-shirts. Suppose that your component choices are choosing a "legs-item" and a "trunk-item." Clearly, there are four options for the legs-item and 12 options for the trunk-item, yielding $4(12) = 48$ possible outfits. Sometimes, of course, you might want to avoid certain combinations, for example a striped shirt with plaid shorts. For this example, suppose you decide that you want to wear either jeans with a t-shirt, or shorts with a tank top. Because you have an equal number of t-shirts and tank tops, for any of the four options for the legs-item, there are six options for the trunk-item, even though the particular options for the trunk-item, t-shirts or tank tops, depends on the option selected on the first choice. Thus, the number of possible outfits is $4(6) = 24$.

Suppose you have four objects, labeled A, B, C, and D, for convenience. A permutation of these four objects is an ordered arrangement of them. For example,

ABCD, DBAC, and ACBD,

are three different permutations of the four objects. Sometimes one is interested in using only some of the objects in forming a permutation. For example,

AB, DB, and AC,

are three different permutations of the four objects A, B, C, and D, *taken two at a time*. Below is a listing of all permutations of the four objects, and a listing of all permutations of the four objects taken two at a time.

ABCD	ABDC	ACBD	ACDB
ADBC	ADCB	BACD	BADC
BCAD	BCDA	BDAC	BDCA
CABD	CADB	CBAD	CBDA
CDAB	CDBA	DABC	DACB
DBAC	DBCA	DCAB	DCBA

AB	AC	AD	BA	BC	BD
CA	CB	CD	DA	DB	DC

Define P_n to be the number of permutations of n objects. From the above listing, $P_4 = 24$. Let $P(n, r)$ be the number of permutations of n objects taken r at a time. From the above listing, $P(4, 2) = 12$. Note that P_n and $P(n, n)$ are the same number.

Instead of the brute force method of listing, the multiplication rule can be used to obtain P_4 and $P(4, 2)$. First, consider $P(4, 2)$. The complex choice is the selection of a permutation. This complex choice has two component choices— the first object in the list, and the second object in the list. There are 4 options for the first object—A, B, C, or D—and for any of these options there are three options for the second object. Thus, by the multiplication rule, $P(4, 2) = 4(3) = 12$.

This method of counting can be extended to the computation of P_4. There are now four component choices—one for each position in the permutation. There are four options for the first position and three options for the second position, as above for $P(4, 2)$. After filling the first two positions, there are two options for the third position. Finally, after filling the first three positions, there is only one option for the final position. Thus,

$$P_4 = 4(3)(2)(1) = 24.$$

Note that the formulas for $P(4, 2)$ and P_4 involve finding the product of consecutive integers; the former is the product of 4 and 3, while the latter is the product of 4, 3, 2, and 1. The following definition will prove to be useful.

> Let r be any positive integer. The number $r!$, read *r-factorial,* is computed as follows.
>
> $$r! = r(r-1)(r-2)\dots 1.$$
>
> In words, begin with r and multiply by successively smaller integers and stop upon reaching 1.
>
> In addition, for convenience define $0! = 1$.

With this definition of r-factorial,

$$P_4 = 4! \text{ and } P(4,2) = 4(3) = 4(3)(\frac{2(1)}{2(1)}) = \frac{4!}{2!}.$$

The argument given above for evaluating $P(4,2)$ and P_4 can be extended to yield the following result.

> Suppose that n and r are positive integers with $r \leq n$. Then
>
> $$P(n,r) = n(n-1)\dots(n-r+1) = \frac{n!}{(n-r)!}.$$
>
> Note that
>
> $$P_n = P(n,n) = \frac{n!}{0!} = n!.$$

Consider again the number of permutations of 4 objects taken 2 at a time, $P(4,2)$. The sequence "AB" is a different permutation than "BA." This fact often is summarized by the phrase, "order matters for permutations." But suppose one does not care about order. Then one is dealing with what are called combinations. For example, listed below are the six different combinations of 4 objects taken 2 at a time.

A, B A, C A, D B, C B, D C, D

Suppose that n and r are positive integers with $r \leq n$. The number of combinations of n objects taken r at a time is denoted by $C(n,r)$. The argument used to obtain a formula for $C(n,r)$ is really quite clever. Suppose you want to compute $P(n,r)$, the number of permutations of n objects taken r at a time. The idea is to take the process of obtaining a permutation and divide it into two steps.

- First, select the r objects that will appear in the permutation.
- Second, take the r objects just selected and permute them.

Next, apply the multiplication rule to these two steps. The result is that $P(n,r)$ equals the product of $C(n,r)$ and P_r:

$$P(n,r) = C(n,r)P_r.$$

Thus,

$$C(n,r) = \frac{P(n,r)}{P_r} = \frac{n!}{(n-r)!r!}.$$

For n any positive integer, it is convenient to define

$$C(n,0) = 1.$$

We are now in a position to compute the number of possible assignments of subjects to treatments in the Infidelity study. The Infidelity study had 20 subjects. An assignment is characterized by identifying the 10 subjects who are assigned to the first treatment (since the remaining 10 subjects necessarily will be assigned to the second treatment). For an assignment order does not matter, so the number of possible assignments for the Infidelity study equals

$$C(20,10) = \frac{20!}{10!10!} = 184{,}756.$$

In general, if a study has n subjects, the number of possible assignments of n_1 subjects to the first treatment and n_2 subjects to the second treatment (where n_1 and n_2 are positive integers with $n_1 + n_2 = n$) equals

$$C(n,n_1) = \frac{n!}{n_1!n_2!}.$$

Remember that if subjects are assigned to treatments by randomization, all assignments are equally likely to occur.

The Sampling Distribution of the Test Statistic for Fisher's Test. Page 62 of the text presents the sampling distribution of the test statistic X for Fisher's test for the Infidelity study. For our present purposes, it is more convenient to work with the number of successes on the first treatment, which we denote by A, as the test statistic. The interested reader can verify the following algebraic relationships between A and X:

$$A = \frac{n_1 n_2 X + n_1 m_1}{n}, \text{ and}$$
$$X = \frac{nA - n_1 m_1}{n_1 n_2}.$$

We will now obtain the sampling distribution of A for the Infidelity study. For convenience, the data from the Infidelity study are reproduced below.

	Response		
Version Read:	Yes	No	Total
1: Tell on Husband?	7	3	10
2: Tell on Wife?	4	6	10
Total	11	9	20

The first task is to determine the possible values of the random variable A. At first glance one might say "Any integer between 0 and 10." This seems reasonable: 10 subjects were assigned to the first treatment so the possible number of successes is any integer between 0 and 10. Remember, however, that the sampling distribution is computed on the assumption that the null hypothesis, or the Skeptic, is correct. This implies that no matter how the subjects are assigned to treatments, there will be 11 successes and nine failures. In other words, if the null hypothesis is true, the marginal totals in the above table are fixed. Since the total number of successes in the study must equal 11, it follows that the random variable A may not take on the value 0 (Why?). Thus, the possible values of A are 1, 2, 3, . . . , 10.

Consider $P(A = 4)$, for example. The event $A = 4$ occurs if, and only if, the contingency table is given by:

	Response		
Version Read:	Yes	No	Total
1: Tell on Husband?	4	6	10
2: Tell on Wife?	7	3	10
Total	11	9	20

We have already determined that the number of possible assignments of subjects to treatments equals

$$C(20, 10) = 184{,}756.$$

To obtain $P(A = 4)$ one must determine the number of assignments that yield the above table. This determination is accomplished by using the multiplication rule. In order to obtain the above table, four of the 11 successes must be assigned to treatment 1; there are $C(11, 4) = 330$ ways to do that. In addition, there are $C(9, 6) = 84$ ways to select six of the nine failures for assignment to treatment 1. Any of the 330 ways to obtain four successes can be combined with any of the 84 ways to obtain six failures, yielding $330(84) =$ 27,721 possible assignments that will yield $A = 4$. Thus,

$$P(A = 4) = \frac{27{,}721}{184{,}756} = 0.1500.$$

The above argument can be extended to other values of a besides four. The result is below.

For the Infidelity study, for $a = 1, 2, \ldots, 10$,

$$P(A = a) = \frac{C(11, a)C(9, 10 - a)}{C(20, 10)}.$$

An irritating feature of the above formula is the need to determine the possible values of A, namely that 0 is not possible as discussed earlier. It turns out that this determination can be avoided with a simple convention. Simply define

$$C(s, t) = 0, \text{ if } t < 0 \text{ or } t > s.$$

With this convention, the above formula for $P(A = a)$ is valid for all nonnegative integers a. (Plugging in $a = 0$ gives $C(9, 10 - a) = C(9, 10) = 0$; any $a > 10$ makes $10 - a < 0$, and, hence, $C(9, 10 - a) = 0$.)

The above argument for the Infidelity study can be extended to any Fisher's test. The result is

For any nonnegative integer a,

$$P(A = a) = \frac{C(m_1, a)C(n - m_1, n_1 - a)}{C(n, n_1)}.$$

These probabilities for A easily yield probabilities for X via the correspondence

$$X = \frac{nA - n_1 m_1}{n_1 n_2},$$

given earlier.