Chapter 16

One Population with Two Dichotomous Responses

This chapter focuses on a new idea. Thus far in these notes, a unit (subject, trial) has yielded one response. In this chapter, we consider situations in which each unit yields two responses, both dichotomies. Later in these Course Notes we will examine situations in which both responses are numbers and the mixed situation of one response being a number and the other a dichotomy. Multi-category responses can be added to the mix, but we won’t have time for this topic.

Sometimes the examples of this chapter will look very much like our examples of Chapter 15. Other times, it will be natural to view our two responses as paired data. As a result, you need to be extra careful as you read through this material.

16.1 Populations: Structure, Notation and Results

A population model for two dichotomous responses can arise for a collection of individuals—a finite population—or as a mathematical model for a process that generates two dichotomous responses per trial.

Here are two examples.

1. Consider the population of students at a small college. The two responses are sex with possible values female and male; and the answer to the following question, with possible values yes and no.

   Do you usually wear corrective lenses when you attend lectures?

2. Recall the data on Larry Bird in Chapter 15, presented in Table 15.16 on page 374. We view his shooting a pair of free throws as a trial with two responses: the outcome of the first shot and the outcome of the second shot.

Recall that I treated the Larry Bird data as Chapter 15 data; i.e., independent random samples from two Bernoulli trials processes. Later in this chapter we will view his results as paired data. Both
perspectives are valid, but it will require some care for you to be comfortable with such *moving between models*. Also, my example of sex and lenses can be viewed as Chapter 15 data, but I would find it awkward to refer to it as paired data.

We begin with some notation. With two responses per unit, sometimes it would be confusing to speak of successes and failures. Instead, we proceed as follows.

- The first response has possible values $A$ and $A^c$. Note that $A^c$ is read *A-complement* or *not-A*.
- The second response has possible values $B$ and $B^c$. Note that $B^c$ is read *B-complement* or *not-B*.

In the above example of a finite population, $A$ could denote female; $A^c$ could denote male; $B$ could denote the answer ‘yes;’ and $B^c$ could denote the answer ‘no.’ In the above example of trials, $A$ could denote that the first shot is made; $A^c$ could denote that the first shot is missed; $B$ could denote that the second shot is made; and $B^c$ could denote that the second shot is missed. In fact, with data naturally viewed as paired, such as Larry Bird’s shots, it is natural to view $A [B]$ as a *success on the first [second] response* and $A^c [B^c]$ as *a failure on the first [second] response*.

It will be easier if we consider finite populations and trials separately. We will begin with finite populations.

### 16.1.1 Finite Populations

Table [16.1] presents our notation for population counts for a finite population. Remember that, in practice, only Nature would know these numbers. This notation is fairly simple to remember: all counts are represented by $N$, with or without subscripts. The symbol $N$ without subscripts represents the total number of members of the population. An $N$ with subscripts counts the number of members of the population with the feature(s) given by the subscripts. For example, $N_{AB}$ is the number of population members with response values $A$ and $B$; $N_{A}$ is the number of population members with value $A^c$ on the first response; i.e., for this, we don’t care about the second response.

Note also that these guys sum in the obvious way:

$$N_A = N_{AB} + N_{AB^c}.$$  

In words, if you take the number of population members whose response values are $A$ and $B$; and add to it the number of population members whose response values are $A$ and $B^c$, then you get the number of population members whose value on the first response is $A$.

It might help if we have some hypothetical values for the population counts. I put some in Table [16.2].

If we take the table of population counts and divide each entry by $N$, we get the table of population proportions or probabilities—see the discussion in the next paragraph. I do this in Tables [16.3] and [16.4] for the general notation and our particular hypothetical data.

Now we must face a notational annoyance. Consider the number 0.36 in Table [16.4] derived from our hypothetical population counts for the sex and lenses study. There are two ways to interpret this number. First, it is the *proportion* of the population who have value $A$ (female) on the
Table 16.1: The table of population counts.

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$N_{AB}$</td>
<td>$N_{AB^c}$</td>
<td>$N_A$</td>
</tr>
<tr>
<td>$A^c$</td>
<td>$N_{A^cB}$</td>
<td>$N_{A^cB^c}$</td>
<td>$N_{A^c}$</td>
</tr>
<tr>
<td>Total</td>
<td>$N_B$</td>
<td>$N_{B^c}$</td>
<td>$N$</td>
</tr>
</tbody>
</table>

Table 16.2: Hypothetical population counts for the study of sex and corrective lenses.

<table>
<thead>
<tr>
<th></th>
<th>Yes ($B$)</th>
<th>No ($B^c$)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female ($A$)</td>
<td>360</td>
<td>240</td>
<td>600</td>
</tr>
<tr>
<td>Male ($A^c$)</td>
<td>140</td>
<td>260</td>
<td>400</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>500</td>
<td>1000</td>
</tr>
</tbody>
</table>

Table 16.3: The table of population proportions—lower case $p$’s with subscripts—or probabilities—upper case $P$’s followed by parentheses.

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$p_{AB} = P(AB)$</td>
<td>$p_{AB^c} = P(AB^c)$</td>
<td>$p_A = P(A)$</td>
</tr>
<tr>
<td>$A^c$</td>
<td>$p_{A^cB} = P(A^cB)$</td>
<td>$p_{A^cB^c} = P(A^cB^c)$</td>
<td>$p_{A^c} = P(A^c)$</td>
</tr>
<tr>
<td>Total</td>
<td>$p_B = P(B)$</td>
<td>$p_{B^c} = P(B^c)$</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 16.4: Hypothetical population proportions or probabilities for the study of sex and corrective lenses.

<table>
<thead>
<tr>
<th></th>
<th>Yes ($B$)</th>
<th>No ($B^c$)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female ($A$)</td>
<td>0.36</td>
<td>0.24</td>
<td>0.60</td>
</tr>
<tr>
<td>Male ($A^c$)</td>
<td>0.14</td>
<td>0.26</td>
<td>0.40</td>
</tr>
<tr>
<td>Total</td>
<td>0.50</td>
<td>0.50</td>
<td>1.00</td>
</tr>
</tbody>
</table>
first response and value B (yes) on the second response. From this perspective, it is natural to view 0.36 as \( p_{AB} \) because we use lower case p’s for population proportions—with a subscript, if needed, to clarify which one. But consider our most commonly used chance mechanism when studying a finite population: Select a member of the population at random. For this chance mechanism it is natural to view 0.36 as the probability of selecting a person who is female and would answer ‘yes.’ We use upper case ‘P’ to denote the word probability. Hence, it is also natural to write \( P(AB) = 0.36 \).

The point of all this is . . . ? Well, in this chapter \( p_{AB} = P(AB) \) (and \( p_A = P(A) \), and so on); the one we use will depend on whether we feel it is more natural to talk about proportions or probabilities.

### 16.1.2 Conditional Probability

Conditional probability allows us to investigate one of the most basic questions in science: How do we make use of partial information?

Consider again the hypothetical population presented in Tables [16.2 and 16.4. Consider the chance mechanism of selecting one person at random from this population. We see that \( P(A) = 0.60 \). In words, the probability is 60% that we will select a female. But suppose we are given the partial information that the person selected answered ‘yes’ to the question. Given this information, what is the probability the person selected is a female? We write this symbolically as \( P(A|B) \); i.e., the probability that \( A \) will occur given that \( B \) occurs. How do we compute it?

We reason as follows. Given that \( B \) occurs, we know that the selected person is among the 500 in column \( B \) of Table [16.2]. Of these 500 persons, reading up the column, we see that 360 are female. Thus, by direct reasoning \( P(A|B) = 360/500 = 0.72 \), which is different than \( P(A) = 0.60 \). In words, knowledge that the person usually wears corrective lenses in lecture increases the probability that the person is female.

We now repeat the above reasoning, but using symbols instead of numbers. Refer to Table [16.1]. Given that \( B \) occurs, we know that the selected subject is among the \( N_B \) subjects in column \( B \). Of these \( N_B \) subjects, reading up the column, we see that \( N_{AB} \) have property \( A \). Thus, by direct reasoning we obtain the following equation.

\[
P(A|B) = \frac{N_{AB}}{N_B}.
\]

(16.1)

Now, this is a perfectly good equation, relating the conditional probability of \( A \) given \( B \) to population counts. Most statisticians, however, prefer a modification of this equation. On the right side of the equation divide both the numerator and denominator by \( N \). This, of course, does not change the value of the right side and has the effect of converting counts to probabilities. The result is below, the equation which is usually referred to as the definition of conditional probability.

\[
P(A|B) = \frac{P(AB)}{P(B)}.
\]

(16.2)

Now there is nothing uniquely special about our wanting to find \( P(A|B) \); we could just as well be interested in, say, \( P(B^c|A^c) \). In fact, there are eight possible conditional probabilities of interest; all combinations of the following three dichotomous choices: to use \( A \) or \( A^c \); to use \( B \) or ..
Table 16.5: Conditional probabilities of the B’s given the A’s in the hypothetical study of sex and lenses. For example, \( P(B|A) = 0.60, P(B^c|A^c) = 0.65 \) and \( P(B^c) = 0.50 \).

<table>
<thead>
<tr>
<th>Yes (B)</th>
<th>No (B^c)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female (A)</td>
<td>0.60</td>
<td>0.40</td>
</tr>
<tr>
<td>Male (A^c)</td>
<td>0.35</td>
<td>0.65</td>
</tr>
<tr>
<td>Unconditional</td>
<td>0.50</td>
<td>0.50</td>
</tr>
</tbody>
</table>

With this interpretation, we immediately know how to compute any conditional probability. For example,

\[
P(B^c|A^c) = \frac{P(A^c B^c)}{P(A^c)}.\]

If you want the conditional probability of one event given another event, calculate the probability that both events occur divided by the probability of the conditioning event occurring.

16.1.3 How Many Probabilities are There?

Look at Table 16.3 again. There are nine probabilities in this table: the four cell probabilities, the four marginal probabilities and the overall probability of 1 in the lower right corner. All except
Table 16.6: Conditional probabilities of the $A$’s given the $B$’s in the hypothetical study of sex and lenses. For example, $P(A|B) = 0.72$, $P(A^c|B^c) = 0.52$ and $P(A^c) = 0.40$.

<table>
<thead>
<tr>
<th></th>
<th>Yes $(B)$</th>
<th>No $(B^c)$</th>
<th>Unconditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female $(A)$</td>
<td>0.72</td>
<td>0.48</td>
<td>0.60</td>
</tr>
<tr>
<td>Male $(A^c)$</td>
<td>0.28</td>
<td>0.52</td>
<td>0.40</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

the 1 are unknown to a researcher, but this does not mean that there are actually eight unknown probabilities. It turns out that if one chooses wisely (remember the very old knight near the end of Indiana Jones and the Last Crusade) then knowledge of three probabilities will suffice to determine all eight probabilities. As we shall see below, there are several possible sets of three, although I won’t give you an exhaustive list of sets. I will give you four of the possible sets that are of the greatest interest to scientists. The first two of my four sets obviously work; the other two require some care and will be covered in a Practice Problem. I will be referring to the symbols in Table 16.3.

1. **Any three of the cell probabilities will suffice.** For example, if we know the values of $P(AB)$, $P(AB^c)$ and $P(A^cB)$, then we can obtain the remaining five unknown probabilities. For example, in Table 16.4 once we know $P(AB) = 0.36$, $P(AB^c) = 0.24$ and $P(A^cB) = 0.14$, we can obtain the remaining probabilities by addition and subtraction.

2. **A row marginal probability, a column marginal probability and any cell probability will suffice.** For example, if we know the values of $P(A)$, $P(B)$ and $P(AB)$, then we can determine all eight probabilities. For example, in Table 16.4 once we know $P(A) = 0.60$, $P(B) = 0.50$ and $P(AB) = 0.36$, we can obtain the remaining probabilities by subtraction and addition.

3. **A conditional probability for each row plus one of the row marginal probabilities will suffice.** For example, if we know the values of $P(B|A)$, $P(B|A^c)$ and $P(A)$, then we can determine all eight probabilities.

4. **A conditional probability for each column plus one of the column marginal probabilities will suffice.** For example, if we know the values of $P(A|B)$, $P(A|B^c)$ and $P(B)$, then we can determine all eight probabilities.

If you choose your three probabilities *unwisely* you will not be able to determine all probabilities. For one of many possible examples, suppose you know $P(A) = 0.80$, $P(B) = 0.30$ and $P(A^c) = 0.20$. With these three probabilities, you cannot determine the remaining five probabilities. (Try it!) The difficulty lies in the fact that once we know $P(A) = 0.80$, we can deduce that $P(A^c) = 0.20$; i.e., knowing $P(A^c)$ is not *new* information.

392
16.1.4 Screening Test for a Disease

You might be thinking, “I am not interested in any relationship between sex and lenses.” Fair enough. I used that example just to get us going. In this subsection we will consider an extremely important application of the ideas of this chapter, namely the analysis of a screening test for a disease.

For many diseases, early detection can be extremely beneficial, for both the ultimate outcome for the patient and the cost of treatment. Screening tests, however, are often controversial. At my annual physical a few years ago, I learned that the PSA screening test for prostate cancer was no longer routine. More recently, there has been much discussion in the media about new recommendations on mammograms for the detection of breast cancer in women.

Here is the way our model fits. We have a population of people at a particular moment in time and we are interested in one particular disease. Each person either has the disease, denoted by $A$, or does not have the disease, denoted by $A^c$. Furthermore, we can imagine giving each person a screening test. For any given person, the screening test can be positive, denoted by $B$, or negative, denoted by $B^c$. Thus, for each person in the population there are two dichotomous responses of interest: the actual disease state and the results of the screening test, if it were given.

We can see immediately that the current problem has issues that were not present in our hypothetical study of sex and lenses. First, it might not be easy to learn whether a person has a disease. (If it were easy, inexpensive and painless to determine, nobody would bother with trying to develop a screening test.) Second, we cannot force a person to have a screening test. (True, we cannot force a person to tell us his/her sex or whether he/she usually wears corrective lenses, but the issue is much trickier for screening tests that might be painful and might have negative consequences.)

As a result, one must use great care in any attempt to evaluate a real life screening test. What I present below is an idealization of what a researcher or physician will face in practice.

Remember that a positive result on a screening test is interpreted as indicating the disease is present. But it is very important to remember that screening tests make mistakes! In a screening test, the various combinations of response values have special meanings and it is important to note these. In particular,

- Event $AB$ is called a correct positive.
- Event $A^c B$ is called a false positive.
- Event $AB^c$ is called a false negative.
- Event $A^c B^c$ is called a correct negative.

Make sure you understand why these labels make sense.

Let’s now look at a hypothetical screening test, presented in Table 16.7. Let’s summarize the information in this table.

1. Ten percent of the population have the disease.

2. If everyone were given the screening test, 18.5% of the population would test positive.
Table 16.7: Population counts for a hypothetical screening test.

<table>
<thead>
<tr>
<th>Screening test result:</th>
<th>Positive (B)</th>
<th>Negative (B')</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Present (A)</td>
<td>95</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>Disease Absent (A')</td>
<td>90</td>
<td>810</td>
<td>900</td>
</tr>
<tr>
<td>Total</td>
<td>185</td>
<td>815</td>
<td>1,000</td>
</tr>
</tbody>
</table>

3. The screening test is correct for 905 persons: 95 correct positives and 810 correct negatives.

4. The screening test is incorrect for 95 persons: 90 false positives and 5 false negatives.

Let’s focus on the errors made by the screening test. (I am definitely a glass half empty person!)

Consider false negatives; i.e., the combination of A, the disease present, with B', the screening test saying the disease is absent. It seems obvious to me that there are three rates, or probabilities, of possible interest. I will illustrate these ideas with our hypothetical screening test.

1. \( P(AB') = 5/1000 = 0.005 \); in words, one-half of one percent of the population will receive a false negative test result.

2. \( P(A|B') = 5/815 = 0.006 \); in words, six-tenths of one percent of those who would test negative will actually have the disease.

3. \( P(B'|A) = 5/100 = 0.05 \); in words, five percent of those who have the disease would test negative.

It seems to me that each of these three numbers—0.005, 0.006 and 0.05 in our example—could reasonably be called a false negative rate. Biostatisticians call \( P(B'|A) \) the false negative rate and, as best I can tell, have not given a name to the other rates.

Does any of this matter? Well, yes, I think it does. First, I would be hard pressed to argue that \( P(AB') \) is an interesting number, but I believe that both of the conditional probabilities are worthy of attention. I believe it is useful to think of \( P(B'|A) \) as being of most interest to the medical community and \( P(A|B') \) as being of most interest to regular people. Why do I say this?

First, consider \( P(B'|A) \), the one that is called the false negative rate by the medical community. As a physician, I might think,

Let’s consider all the persons with the disease; what proportion of these people who need help will be told that they are fine?

Second, consider \( P(A|B') \), the one with no name. A person is told that the screening test result is negative. A thoughtful person—some would say hypochondriac, but let’s not be judgmental—might wonder,

Do I have the disease; i.e., did the screening test make a mistake?

394
A definitive answer might be available only with an autopsy—thanks, but I’ll pass on that! There is some value, however, in considering the number \( P(A|B^c) \). Of all the persons who test negative, this number is the proportion that actually have the disease.

As an aside, I hope that you have not concluded that I view the medical research community as some nasty organization that gives names only to things of interest to them and refuses to name things of interest to patients. As we will see later in this chapter, data collection might well give us a good estimate of \( P(B^c|A) \), but it is often impossible to estimate—without some controversial assumptions—the value of \( P(A|B^c) \). My suspicion is that \( P(B^c|A) \) is given a name, in part, because it can be estimated without controversy. But, of course, I could be wrong.

My issue of there being two rates of interest becomes really important when we consider false positives. A false positive occurs whenever \( A^c \) is matched with \( B \). Again, there are three rates we could calculate, again illustrated with our hypothetical screening test.

1. \( P(A^cB) = 90/1000 = 0.09 \); in words, nine percent of the population will receive a false positive test result.

2. \( P(B|A^c) = 90/900 = 0.10 \); in words, ten percent of those for whom the disease is absent would test positive.

3. \( P(A^c|B) = 90/185 = 0.487 \); in words, 48.7% of those who would test positive are free of the disease.

Look at these last two rates. Biostatisticians call \( P(B|A^c) \) the false positive rate, but it seems to me that \( P(A^c|B) \) deserves a name too! At various times in history, governments have considered or enacted policies in which: everybody gets tested and those who test positive will be quarantined. This last computation shows that, for our hypothetical population, of those quarantined, 48.7% are actually disease free!

### 16.1.5 Trials and Bayes’ Formula

We do not have a table of population counts for trials. We begin with the table of probabilities.

So, how do we obtain the table of probabilities? Well, mostly, we cannot obtain it, but in later sections we will learn how to estimate it from data. But this is a good time to introduce a couple of important ideas.

Recall the definition of conditional probability given in Equation [16.2]

\[
P(A|B) = P(AB)/P(B).
\]

We can rewrite this as

\[
P(AB) = P(B)P(A|B).
\]

This new equation is called the multiplication rule for conditional probabilities. Note what it says, in words:

The probability that two events both occur is the product of the probability of one event occurring and the conditional probability of the remaining event, given the one we already handled.

395
With this admittedly awkward statement, we can obtain two versions of the multiplication rule for conditional probabilities. For example, if we interchange the roles of $A$ and $B$ in Equation 16.3 we get:

$$P(AB) = P(A)P(B|A).$$  \hspace{1cm} (16.4)$$

We want two ways to calculate $P(AB)$ because sometimes we know the pair $P(A)$ and $P(B|A)$ and other times we know the pair $P(B)$ and $P(A|B)$.

I will now give you a somewhat silly application of these ideas to illustrate how we can build a table of probabilities.

Years ago, I lived with a dog named Casey. Periodically, I would exercise on my treadmill for 30 minutes and watch television. Casey would sit by the window watching the yard, with special interest in viewing squirrels. I could not see the yard from the treadmill.

I will view each such 30 minute segment of time as a trial. Two dichotomous responses are obtained—though not by me—during the trial:

- $A$: One or more squirrels enter the yard; obviously, $A^c$ is that no squirrels enter the yard.
- $B$: Casey barks at some time during the trial; obviously, $B^c$ is that Casey does not bark during the trial.

From past experience I know (estimate or guess might be more accurate verbs; see the next section) the following numbers:

$$P(A) = 0.30, \ P(B|A) = 0.80 \text{ and } P(B|A^c) = 0.10.$$ 

In words, in any given trial, there is a 30% chance that at least one squirrel will visit the yard; given that at least one squirrel visits the yard, there is an 80% chance that Casey will bark; and if no squirrels visit the yard, there is a 10% chance that Casey will bark.

The first fact we can deduce is: $P(A^c) = 1 - P(A) = 0.70$. We now proceed to complete the following table.

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td></td>
<td></td>
<td>0.30</td>
</tr>
<tr>
<td>$A^c$</td>
<td></td>
<td></td>
<td>0.70</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

Next, we consider $P(AB)$. By Equation 16.3 $P(AB) = P(B)P(A|B)$, but this is no help, because we know neither of the numbers on the right side of this equation. Instead, we use Equation 16.4.

$$P(AB) = P(A)P(B|A) = 0.30(0.80) = 0.24.$$ 

Similarly,

$$P(A^cB) = P(A^c)P(B|A^c) = 0.70(0.10) = 0.07.$$ 

Next, we place these two numbers in our table and continue with simple subtractions and additions until we obtain the completed table, given below.
There are some amazing facts revealed in this table! First, we see that \( P(B) = 0.31 \); i.e., there is a 31\% chance that Casey will bark during a trial. Why is this amazing? Well, we started with information on whether Casey would bark conditional on squirrel behavior, and end up with the unconditional probability of Casey barking.

OK, well maybe the former was not so amazing, but this next one definitely is. It is so great that it has a name: Bayes’ formula or Bayes’ rule; well, two names. It is named in honor of the Reverend Thomas Bayes who did or did not discover it before his death in 1761. (The historical controversy will be ignored in this course.)

First, I will show you how to use Bayes’ rule, which is very easy, and then I will give you the formula, which is quite a mess. Suppose that during my trial I hear Casey bark. It is natural for me to wonder, “Did any squirrels visit the yard?” In other words, I want to calculate \( P(A|B) \).

Now before Bayes nobody could answer this question; in fact, it looked impossible: we are given information about conditional probabilities of \( B \)’s given \( A \)’s, how could we possibly reverse them? It seemed like alchemy or some occult method would be required to obtain an answer.

But as often happens, especially in math or riding a bicycle, what appears to be illogical or impossible works if you just do it. (Ugh! This sounds like a Nike commercial; what’s next? Impossible is nothing?)

Let’s just calculate \( P(A|B) \); but how? Well, by definition, \( P(A|B) = P(AB)/P(B) \) and we can read both of these numbers from our table! Thus,

\[
P(A|B) = \frac{P(AB)}{P(B)} = \frac{0.24}{0.31} = 0.774.
\]

In words, given that Casey barks, there is a 77.4\% chance that at least one squirrel visited the yard.

We see that it is easy to reverse the direction of conditioning, provided we are able to complete the table of probabilities. My advice is that in practice, just complete the table and then you can calculate any conditional probability that you desire.

For completeness, I will show you Bayes’ formula in all its mathematical glory, but I do not recommend using it and you are not responsible for it in any way!

Here are elements we need for Bayes’ formula.

- We need a partition of the sample space, denoted by the \( k \) events \( A_1, A_2, \ldots, A_k \). By a partition I mean that the events are pairwise disjoint (i.e., they don’t overlap in any way) and their union is the entire sample space.

- We need some other event of interest \( B \).

- We need to know \( P(A_i) \), for \( i = 1, 2, \ldots, k \).

- We need to know \( P(B|A_i) \), for \( i = 1, 2, \ldots, k \).
Before I give you Bayes’ formula note that these three conditions are met for my example with Casey: \( k = 2 \); the events are \( A_1 = A \) and \( A_2 = A^c \); \( B \) is as above; and both probabilities and both conditional probabilities are known.

So, here it is, Bayes’ formula:

\[
P(A_i | B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{k} P(A_j)P(B|A_j)}, \text{ for } i = 1, 2, \ldots, k.
\] (16.5)

## 16.2 Random Samples from a Finite Population

Recall our table of probabilities, reproduced below:

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>( P(AB) )</td>
<td>( P(AB^c) )</td>
<td>( P(A) )</td>
</tr>
<tr>
<td>( A^c )</td>
<td>( P(A^cB) )</td>
<td>( P(A^cB^c) )</td>
<td>( P(A^c) )</td>
</tr>
<tr>
<td>Total</td>
<td>( P(B) )</td>
<td>( P(B^c) )</td>
<td>1</td>
</tr>
</tbody>
</table>

In my experience, in most scientific applications these eight probabilities are unknown. Also, a scientist might well be interested in any or all of the eight possible conditional probabilities. (Admittedly, as discussed earlier, there are many mathematical relationships between these 16 unknown rates.) In this section we will discuss what is possible and what is practicable in estimating these various numbers.

We will investigate three types of random samples. Whichever way we sample, we will present the data we obtain in a table, as illustrated in Table 16.8; our ubiquitous \(2 \times 2\) table for data, introduced in Chapter 15 in Table 15.1 on page 354. I introduce the three types of random samples below and I will illustrate each with our earlier example of sex and lenses.

1. **Type 1: Overall Random Sample.** Select a random sample of size \( n \) from the population. In this case all eight of the remaining numbers in Table 16.8 (excluding the \( n \)) are the observed values of random variables; i.e., their values cannot be predicted, with certainty, before data collection.

2. **Type 2: Independent Random Samples from the Rows.** Select two independent random samples: The first is of size \( n_1 \) from the population of all subjects with value \( A \) for the first response; the second is of size \( n_2 \) from the population of all subjects with value \( A^c \) for the first response. In this case, \( n_1, n_2 \) and \( n = n_1 + n_2 \) are all fixed in advance by the researcher; the remaining six counts are observed values of random variables.
3. **Type 3: Independent Random Samples from the Columns.** Select two independent random samples: The first is of size $m_1$ from the population of all subjects with value $B$ for the second response; the second is of size $m_2$ from the population of all subjects with value $B^c$ for the second response. In this case, $m_1$, $m_2$ and $n = m_1 + m_2$ are all fixed in advance by the researcher; the remaining six counts are observed values of random variables.

For our sex and lenses study, these become:

1. **Type 1: Overall Random Sample.** A random sample of size $n$ is selected from the population of all 1,000 students.

2. **Type 2: Independent Random Samples from the Rows.** Two lists are created: one of the 600 female students and one of the 400 male students. Select a random sample of size $n_1$ from the female population. Then select an independent random sample of size $n_2$ from the male population.

3. **Type 3: Independent Random Samples from the Columns.** Two lists are created: one of the 500 students who would answer ‘yes’ and one of the 500 students who would answer ‘no.’ Select a random sample of size $m_1$ from the ‘yes’ population. Then select an independent random sample of size $m_2$ from the ‘no’ population.

Note that it is often the case that at least one of the three ways of sampling is unrealistic. In the above, I cannot imagine that the researcher would have a list of either the ‘yes’ or ‘no’ population; hence, the Type 3 sampling is not of interest for this example.

Let us consider Type 2 sampling in general. Upon reflection, you will realize that Type 2 sampling is equivalent to what we studied in Chapter 15, except that the names have been changed. In particular, population 1 in Chapter 15 consists of all subjects with feature $A$ and population 2 in Chapter 15 consists of all subjects with feature $A^c$. In this context, label $B$ a success and $B^c$ a failure. Thus, what we earlier called $p_1$ and $p_2$ are now $P(B|A)$ and $P(B|A^c)$, respectively. Thus, our earlier methods for estimation and testing of $p_1 - p_2$ can be immediately applied to a difference of conditional probabilities: $P(B|A) - P(B|A^c)$.

Of course, Type 2 and Type 3 sampling are really the same mathematically. For example, if you have independent random samples from the columns, you may simply interchange the roles of rows and columns and then have independent random samples from the rows. In practice it is convenient to allow for both Type 2 and Type 3 sampling.

For independent random samples from the columns, identify population $1^*$ as all subjects with feature $B$ and population $2^*$ as all subjects with feature $B^c$. Make the definitions $p_1^* = P(A|B)$ and $p_2^* = P(A|B^c)$. Thus, the difference in conditional probabilities $P(A|B) - P(A|B^c)$ is simply $p_1^* - p_2^*$.

The above might seem very confusing, but it’s really quite simple: Sample by rows and we are interested in $p_1 - p_2$; sample by columns and we are interested in $p_1^* - p_2^*$.

There are six parameters of most interest to a researcher:

\[ p_A, p_B, p_1, p_2, p_1^*, p_2^*. \]

Now, here are the really important facts to note:
1. For Type 1 sampling, all six of these parameters can be estimated.

2. For Type 2 sampling, only $p_1$ and $p_2$ can be estimated; the other four cannot be estimated.

3. For Type 3 sampling, only $p_1^*$ and $p_2^*$ can be estimated; the other four cannot be estimated.

We can see the truth of these facts with a simple, but extreme, example. In the sex and lenses study, suppose that I decide to take a Type 2 sample as follows: $n_1 = 10$ of the 600 females and all 400 males. Then I will get the following data:

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$a$</td>
<td>$10 - a$</td>
<td>10</td>
</tr>
<tr>
<td>$A^c$</td>
<td>140</td>
<td>260</td>
<td>400</td>
</tr>
<tr>
<td>Total</td>
<td>$140 + a$</td>
<td>$270 - a$</td>
<td>410</td>
</tr>
</tbody>
</table>

We know from the population counts that $P(A|B) = p_1^* = 360/500 = 0.72$. But our estimate of $p_1^*$ from this table will be $a/(140 + a)$ which can range from a minimum of 0 when $a = 0$ to a maximum of 0.067 when $a = 10$. In other words, our estimate of 0.72 will never be larger than 0.067! **This is a very bad estimate!** Think about this example and make sure that you understand why it is happening.

Before leaving this section, I want to revisit the screening test example in the context of our three types of sampling.

Medical researchers typically do not use Type 1 sampling for the following two reasons.

1. Many diseases are quite rare, making $P(A)$, $P(AB)$ and $P(AB^c)$ very small. As a result, one would require a huge sample size to estimate these well. In the best of situations a huge sample size is expensive. Also, nobody would spend a huge amount of money on sampling just to (possibly) learn that the screening test is ineffective.

2. We skirt around the issue of how difficult it is to obtain a random sample. For many diseases, the sufferers are people who are not particularly easy to find, which, of course, is an important step in being in a sample. For example, IV drug users and sex workers are two groups that have high rates of HIV infection, but are difficult to locate for a study. And, if located, they might be reluctant to participate.

Type 3 sampling is not possible because researchers will not have lists of people who will test positive (negative) before they collect data!

As a result, the most realistic way to sample is Type 2: Select what you hope is a random sample from people who clearly have the disease and an independent random sample from people who seem to not have the disease and proceed.

Thus, confirming what I stated earlier, because they use Type 2 sampling, medical researchers can estimate $p_1$ and $p_2$. They cannot estimate either $p_1^*$ or $p_2^*$. Thus, it is understandable why they refer to the former with the definite article ‘the.’
Table 16.9: Hypothetical population counts, in thousands.

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Outcome</th>
<th>Bad (B)</th>
<th>Good (B')</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present (A)</td>
<td>24</td>
<td>276</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Absent (A')</td>
<td>28</td>
<td>672</td>
<td>700</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>52</td>
<td>948</td>
<td>1,000</td>
<td></td>
</tr>
</tbody>
</table>

16.3 Relative Risk and the Odds Ratio

Let’s return to the model of a finite population with two responses, as introduced at the beginning of this chapter. The following situation is common in medical studies. The first response is the presence (A) or absence (A') of a risk factor. The second response is a health outcome, either bad (B) or good (B'). Here is an example. A pregnant woman is a smoker (A) or nonsmoker (A'). Her baby’s birth-weight is either low (B) or normal (B').

We will begin with a table of hypothetical population counts, presented in Table 16.9. Typically, a researcher is interested in comparing $P(B|A)$ to $P(B|A')$. These are the probabilities of the bad outcome conditional on the risk feature being present or absent, respectively. Even though B is an undesirable outcome, we label it a success because in these medical problems it is often rare. Even if it’s not rare in a particular problem, we still call it a success to avoid confusion. (Having B sometimes be a success and sometimes be a failure would definitely confuse me!) Using Chapter 15 notation, I write $p_1 = P(B|A)$ and $p_2 = P(B|A')$.

We note that for the hypothetical population counts in Table 16.9, $p_1 = 24/300 = 0.08$ and $p_2 = 28/700 = 0.04$. We could compare these by subtracting:

$$p_1 - p_2 = 0.08 - 0.04 = 0.04,$$

in words, the probability of the bad outcome is larger by 0.04 when the risk factor is present compared to when the risk factor is absent. Because both $p_1$ and $p_2$ are small, we might want to compare them by dividing: $p_1/p_2$. This ratio is called the relative risk.

For the population counts in Table 16.9, the relative risk equals $p_1/p_2 = 0.08/0.04 = 2$. In words, the presence of the risk factor doubles the probability of the bad outcome.

Scientifically, it would make sense to want to estimate $p_1 - p_2$ or $p_1/p_2$. Recall the three types of random sampling that were introduced on page 398. Type 1 sampling is rarely used because it is difficult to get a random sample from the entire population and even if we could obtain one, with a rare bad outcome we won’t get enough data for subjects with bad outcomes to learn very much. Also, it can be difficult to perform Type 2 sampling, because there won’t be lists of people based on the presence or absence of the risk factor. Type 3 sampling is, however, often reasonable. One can use hospital records to obtain—one hopes—Type 3 random samples. In fact, a study based on Type 3 sampling is called a case-control study. As we have argued earlier, however, with Type 3 sampling we cannot estimate $p_1$ or $p_2$. We can handle this difficulty, somewhat, by introducing the notion of the odds ratio.
Odds are an alternative to probabilities as a measure of uncertainty. For example, consider one cast of a balanced die and suppose we are interested in the outcome ‘4.’ We have said that the probability of the outcome ‘4’ is $\frac{1}{6}$, but we could also say that the odds of the ‘4’ are 1 (the number of outcomes that give ‘4’) to 5 (the number of outcomes that don’t give ‘4’), or $\frac{1}{5}$. In general, if the event of interest has probability $p$, then the odds of the event is $\frac{p}{1 - p}$.

The current paragraph is enrichment. I recommend you read it, but you will not be responsible for its content. Be careful with language! I am talking about the odds of an event. Sometimes people—especially in gambling contexts—speak of the odds against an event, which is the inverse of the odds of the event. Thus, a gambler would say that, when casting a balanced die, the odds against a ‘4’ are the inverse of $\frac{1}{5}$, which is 5, usually stated as 5 to 1. But gambling is actually more complicated because bookies and casinos modify the true odds against an event to ensure themselves a long-run profit. We saw this earlier with the roulette wheel example. The probability of red is $\frac{18}{38}$. Remembering that the odds of an event with probability $p$ is $\frac{p}{1 - p}$, the odds against the event is $\frac{1 - p}{p}$. Thus, the odds against red are $\frac{20}{18} = \frac{10}{9}$. It would be a fair bet if the casino paid $19$ for a $9$ bet that wins. Casinos, of course, have no desire to offer fair bets. An easy way for them to deal with this is to take the actual odds against, $\frac{10}{9}$ for the roulette bet, and shorten (reduce) the odds; in the case of roulette they shorten the odds against from the true $\frac{10}{9}$ to the profitable 1.

For the problem of this section, the odds of $B$ in row $A$ is $p_1/(1 - p_1)$ and the odds of $B$ in row $A^c$ is $p_2/(1 - p_2)$. In terms of the symbols in Table [16.1] these odds are $N_{AB}/N_{AB^c}$ and $N_{A^cB}/N_{A^cB^c}$, respectively. Thus, their ratio, called the odds ratio, is:

$$\frac{N_{AB}N_{A^cB^c}}{N_{AB^c}N_{A^cB}}.$$  \hfill (16.6)

The great thing about this formula is the following. We have defined the odds ratio in terms of the row probabilities; i.e., the conditional probabilities of the $B$’s given the $A$’s. But an examination of this formula shows that it is symmetric in the arguments $A$ and $B$; hence, the odds ratio remains the same if we define it in terms of the column probabilities. Thus, and this is the important part, we can estimate the odds ratio for any of our three types of sampling.

Finally, a little bit of algebra shows that

$$\text{odds ratio} = \text{relative risk} \times \frac{1 - p_2}{1 - p_1}.$$  

Thus, if both $p_1$ and $p_2$ are close to zero, the odds ratio is approximately equal to the relative risk; and we know that the relative risk is interesting.

For the population in Table [16.9] the odds ratio is

$$\frac{24(672)}{276(28)} = 2.087,$$

which is close to the relative risk, previously shown to equal 2.

We will now discuss how to estimate the odds ratio. Note that this method is valid for any of our three types of random sampling. Table [16.10] presents our notation for the data we collect, our
Table 16.10: Notation for data for estimating the odds ratio.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Risk Factor</th>
<th>Bad ($B$)</th>
<th>Good ($B^c$)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present ($A$)</td>
<td>$a$</td>
<td>$b$</td>
<td>$n_1$</td>
<td></td>
</tr>
<tr>
<td>Absent ($A^c$)</td>
<td>$c$</td>
<td>$d$</td>
<td>$n_2$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$m_1$</td>
<td>$m_2$</td>
<td>$n$</td>
<td></td>
</tr>
</tbody>
</table>

ubiquitous $2 \times 2$ table of data in Table 15.1. Our point estimate of the odds ratio, $\theta$, is

$$\hat{\theta} = \frac{ad}{bc}.$$  

I placed the population in Table 16.9 in my computer and simulated a case-control study with $m_1 = m_2 = 200$. My data are in Table 16.11. My estimated odds ratio is

$$\hat{\theta} = \frac{[89(131)]}{[69(111)]} = 1.522,$$

which is considerably smaller than the population value, which Nature alone knows to be 2.087.

We can obtain a confidence interval estimate of $\theta$ but it’s a bit involved. We actually obtain a confidence interval estimate of $\lambda = \ln(\theta)$, where by ‘ln’ I mean the natural logarithm that is popular in calculus and has base $e = 2.71828 \ldots$.

Our point estimate of $\lambda$ is $\lambda = \ln(\hat{\theta})$. For our data, $\hat{\lambda} = \ln(1.522) = 0.4200$.

The approximate confidence interval estimate of $\lambda$ is:

$$\hat{\lambda} \pm z^* \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}. \tag{16.7}$$

The 95% confidence interval estimate of $\lambda$ for the data in Table 16.11 is:

$$0.4200 \pm 1.96 \sqrt{\frac{1}{89} + \frac{1}{69} + \frac{1}{111} + \frac{1}{131}} =$$

$$0.4200 \pm 1.96(0.2058) = 0.4200 \pm 0.4035 = [0.0165, 0.8235].$$

To get back to a confidence interval estimate of $\theta$, we exponentiate the endpoints of this interval:

$$(2.71828)^{0.0165} = 1.017 \text{ and } (2.71828)^{0.8235} = 2.278.$$  

Thus, finally, the 95% confidence interval estimate of the odds ratio is $[1.017, 2.278]$. This is not a very useful interval; it is correct, because it contains $\theta = 2.087$, but its lower bound is so close to 1, the risk factor being present might not be much of a danger. (An odds ratio of 1 means that $p_1 = p_2$.) More data are needed.
Table 16.11: Simulated data for estimating the odds ratio from the population given in Table 16.9

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Bad ($B$)</th>
<th>Good ($B^c$)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present ($A$)</td>
<td>89</td>
<td>69</td>
<td>158</td>
</tr>
<tr>
<td>Absent ($A^c$)</td>
<td>111</td>
<td>131</td>
<td>242</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
</tbody>
</table>

16.4 Comparing $P(A)$ to $P(B)$

For a finite population, this section assumes we have a Type 1 random sample; i.e., a random sample from the entire population. For trials, the assumption is that we have i.i.d. dichotomous trials; i.e., the table of probabilities is the same for each trial and trials are independent. This does not mean that $A$ and $B$ are independent—see Practice Problem 2.

In this section, I will focus on the entries $P(A)$ and $P(B)$ in Table 16.3 on page 389. In many studies, it would be difficult to explain an interest in comparing these two probabilities. In my hypothetical study of sex and lenses, why would I want to compare the proportion of females in the population ($P(A)$) with the proportion of people who would answer the question with ‘yes’ ($P(B)$)? Similar comments are true for the studies: of a screening test; my dog barking at squirrels; the relationship between a risk factor and a bad outcome.

In my study of Larry Bird, however, $P(A)$ is the probability that he makes the first of the pair of free throws and $P(B)$ is the probability that he makes the second of the pair of free throws. As a result, it is natural to compare $P(A)$ and $P(B)$ to investigate whether he is more skilled on his first or second shot.

I will not begin this section, however, with this basketball example. Instead we will revisit a technique we used extensively in Part I of these Course Notes: computer simulation experiments.

16.4.1 The Computer Simulation of Power

Refer to Table 9.8 on page 203. This table presents the results of a power study for a variety of constant treatment effect alternatives for Sara’s golf study. It should not be a problem if your recollection of this topic and this table are a bit hazy. My point is more about simulation experiments and not really so much about power.

Recall that I stated that the number of possible assignments for Sara’s golf study is $1.075 \times 10^{23}$. I am interested in the finite population that consists of

$$N = 1.075 \times 10^{23}$$

members, with each member being a possible assignment. Each member’s card contains two dichotomous pieces of information:

- $A$ if test statistic $U$ would reject the null hypothesis for the card’s assignment; and $A^c$ if test statistic $U$ would fail to reject the null hypothesis for the card’s assignment.
Table 16.12: Results of a computer simulation experiment on the power of test statistics $U$ and $R_1$ for the alternative of a constant treatment effect of 3 yards for Sara’s golf study.

<table>
<thead>
<tr>
<th></th>
<th>$U$</th>
<th>$R_1$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject ($A$)</td>
<td>1,136</td>
<td>153</td>
<td>1,289</td>
</tr>
<tr>
<td>Fail to reject ($A^c$)</td>
<td>398</td>
<td>8,313</td>
<td>8,711</td>
</tr>
<tr>
<td>Total</td>
<td>1,534</td>
<td>8,466</td>
<td>10,000</td>
</tr>
</tbody>
</table>

- $B$ if test statistic $R_1$ would reject the null hypothesis for the card’s assignment; and $B^c$ if test statistic $R_1$ would fail to reject the null hypothesis for the card’s assignment.

Recall from Table 9.7 on page 201, that the critical regions for these tests are $U \geq 10.65$ and $R_1 \geq 1788.0$. The critical values—10.65 and 1788.0—were obtained via a computer simulation experiment with 10,000 reps and both give, approximately, $\alpha = 0.0499$ as the probability of a Type I error.

In Table 9.8 on page 203, I report the results of six computer simulation experiments, one each for six hypothesized values of the constant treatment effect. Thus, I have data for six different population boxes; one box for each of the effects studied. I will restrict attention for now to a constant treatment effect of 3 yards.

With the above set-up, $P(A)$ is the probability that test statistic $U$ would correctly reject the null hypothesis given that the constant treatment effect equals 3 yards. Similarly, $P(B)$ is the probability that test statistic $R_1$ would correctly reject the null hypothesis given that the constant treatment effect equals 3 yards. If we could examine all $1.075 \times 10^{23}$ possible assignments, then we would know the values of $P(A)$ and $P(B)$ and, hence, be able to determine which test statistic is more powerful for the alternative we are studying. Of course, we might learn that $P(A) = P(B)$, which would tell us that the tests are equally powerful for the given alternative.

Instead of examining all possible assignments, I used a computer simulation experiment with 10,000 reps; in the terminology introduced in Chapter 10, I selected a dumb random sample of size $n = 10,000$.

I begin by reprinting the results of my computer simulation experiment in Table 16.12. The careful reader will note that I have interchanged the rows and columns in my Chapter 9 table to obtain this Chapter 16 table, so that, in the current chapter’s set-up $P(A)$ and $P(B)$ refer to power. I will now show you how to use the data in Table 16.12 to make an inference about the values of $P(A)$ and $P(B)$.

Obviously, we can consider these probabilities separately. In particular, focusing on test statistic $U$, we find that our point estimate of $P(A)$ is 0.1289 and the approximate 95% confidence interval estimate of $P(A)$ is:

$$0.1289 \pm 1.96 \sqrt{0.1289(0.8711)/10,000} = 0.1289 \pm 0.0066 = [0.1223, 0.1355].$$

Similarly, our point estimate of $P(B)$ is 0.1534 and the approximate 95% confidence interval
estimate of $P(B)$ is:

$$0.1534 \pm 1.96 \sqrt{0.1534(0.8466)/10,000} = 0.1534 \pm 0.0071 = [0.1463, 0.1605].$$

These intervals are comfortably separated; the upper bound of the former is 0.0108 smaller than
the lower bound of the latter. Thus, it seems clear that $R_1$ is more powerful than $U$.

Our confidence interval formula for comparing two proportions requires us to have independent
random samples. What would this entail for the current problem? We would select a dumb random
sample of 10,000 assignments to evaluate the performance of $U$; then we would again select a
dumb random sample of 10,000 assignments to evaluate the performance of $R_1$. In other words,
we would perform a computer simulation experiment with 10,000 reps for $U$ and then perform a
computer simulation experiment with 10,000 reps for $R_1$. I could have done this; why didn’t I?

Well, let’s pretend that I had performed two computer simulation experiments and again
obtained point estimates of 0.1289 and 0.1534. Let’s find the 95% confidence interval estimate of
$P(A) - P(B)$:

$$\left(0.1289 - 0.1534\right) \pm 1.96 \sqrt{\frac{0.1289(0.8711)}{10,000} + \frac{0.1534(0.8466)}{10,000}} =$$

$$-0.0245 \pm 0.0096 = [-0.0341, -0.0249].$$

Please note the half-width of this interval: 0.0096. When we learn the correct confidence interval
estimate later in this section, we will see that its half-width is only about one-half as much, 0.0046.
Thus, by having one sample of assignments and using each assignment twice, we reduce the half-
width, in this example, by a factor of two when compared to having two independent samples.

Let’s now learn the correct way to analyze the data in Table 16.12. We will begin with a test of
hypotheses.

Based on Occam’s Razor, the natural null hypothesis is $P(A) = P(B)$. There are three natural
possibilities for the alternative hypothesis:

$H_1$: $P(A) > P(B)$; or $H_1$: $P(A) < P(B)$; or $H_1$: $P(A) \neq P(B)$.

Using the symbols in Table 16.3 on page 389

$P(A) = P(B)$ becomes $P(AB) + P(AB^c) = P(AB) + P(A^cB)$,

which, after canceling, becomes $P(AB^c) = P(A^cB)$.

Similarly, the alternative $>$ is equivalent to

$P(AB^c) > P(A^cB)$;

the alternative $<$ is equivalent to

$P(AB^c) < P(A^cB)$;

and the alternative $\neq$ is equivalent to

$P(AB^c) \neq P(A^cB)$. 

406
Below is a partial reproduction of the table of probabilities:

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>B^c</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>P(AB^c)</td>
<td></td>
</tr>
<tr>
<td>A^c</td>
<td>P(A^cB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The two probabilities in this picture are the only ones that matter for the test of hypotheses.

This is a very interesting result. Of the four cells in the population table, only two are relevant to our null hypothesis.

Let’s look at the data in Table 16.12. In this table, we have the results of a dumb random sample of size 10,000, but the numbers \( a = 1,136 \) and \( d = 8,313 \) are irrelevant for our test of hypotheses. What are relevant are the numbers \( b = 153 \) and \( c = 398 \).

Define \( m = b + c \) which equals \( 153 + 398 = 551 \) for our data. I call \( m \) the effective sample size. The sample size is 10,000, but of these observations, only \( m = 551 \) are relevant for our test. We can think of \( m \) as the observed value of the random variable \( M \); before collecting data, we define \( M \) to be the number of observations that fall into either of the cells on the off diagonal: \( AB^c \) or \( A^cB \).

This next part is a mess notationally, but a simple idea. Conditional on the knowledge that an observation falls in an off diagonal cell, label the upper right cell—\( AB^c \)—a success and the lower left cell—\( A^cB \)—a failure. Thus, conditional on the value \( M = m \), we have \( m \) Bernoulli trials. For these \( m \) Bernoulli trials, the probability of success is

\[
P(AB^c | AB^c \text{ or } A^cB) = \frac{P(AB^c)}{P(AB^c) + P(A^cB)} \text{ which we will call } p.
\]

Now here is the key point to note.

- If the null hypothesis is true, then \( p = 0.5 \).
- If the alternative hypothesis \( > \) is true, then \( p > 0.5 \).
- If the alternative hypothesis \( < \) is true, then \( p < 0.5 \).
- If the alternative hypothesis \( \neq \) is true, then \( p \neq 0.5 \).

Thus, our test of hypotheses can be viewed as testing

\[
H_0: p = 0.5 \text{ versus } H_1: p > 0.5; \text{ or } H_1: p < 0.5; \text{ or } H_1: p \neq 0.5,
\]

where \( p \) is the probability of success for \( m \) Bernoulli trials. **We already know how to perform this test!** Its P-value is given by Result 12.2 on page 306 with \( p_0 = 0.5 \).

Because our special value of interest is 0.5, the the null sampling distribution is \( \text{Bin}(m, 0.5) \), which is symmetric. As a result, the two terms that are summed for the alternative \( \neq \) are the same number.

In the context of this chapter, the test from Chapter 12 is called McNemar’s test. I will use our data on power to remind you how to use the website:
to obtain the exact P-value for McNemar’s test.

After accessing the site, enter 0.5 as the **Probability of success**; enter \( m = 551 \) for the **Number of trials**; and enter \( b = 153 \) for the **Number of successes**. Click on calculate and you will obtain five probabilities. The two that are relevant are:

- \( P(X \leq 153) = 1.920 \times 10^{-26} \); and
- \( P(X \geq 153) = 1 \).

The first of these is the exact P-value for the alternative \(<\>; the second of these is the exact P-value for the alternative \(>\); and twice the first of these is the exact P-value for the alternative \(\neq\). I should have mentioned it earlier, but the only defensible choice for the alternative for the power study is \(\neq\). If one is trying to decide which test statistic is more powerful, why would one ever use a one-sided alternative!

Thus, the exact P-value for my alternative is \(3.840 \times 10^{-26}\). **This is a very small P-value!** Could it be an error? Well, the mean and standard deviation of the Bin(551,0.5) distribution are

\[
\mu = 551(0.5) = 275.5 \quad \text{and} \quad \sigma = \sqrt{551(0.5)(0.5)} = 11.73.
\]

Thus, the observed value \( b = 153 \) is more than ten standard deviations below the mean! As a statistician I tend to be cautious; but not in this problem! I know that the sum of ranks test is more powerful than the comparison of means test for the alternative we have considered.

We went to a great deal of effort to motivate the test of hypotheses for comparing \(P(A)\) and \(P(B)\); for a change-of-pace, I simply will give you the approximate confidence interval estimate of \(P(A) - P(B)\). (Actually, the formula below is a special case of a formula you will learn in Chapter 20.) The approximate confidence interval estimate of \(P(A) - P(B)\) is:

\[
\frac{b - c}{n} \pm \frac{z^*/n}{\sqrt{\frac{n(b + c) - (b - c)^2}{n - 1}}},
\]  
(16.8)

where, as usual, the value of \(z^*\) depends on one’s choice of the confidence level, as given in Table 12.1. I will illustrate the use of this formula for the data of our study of power. Recall that \(b = 153\), \(c = 398\) and \(n = 10,000\). First, \((b - c) = -245\); \((b + c) = 551\); and

\[
n(b + c) - (b - c)^2 = 10000(551) - (-245)^2 = 5,449,975.
\]

Thus, the approximate 95% confidence interval estimate of \(P(A) - P(B)\) is:

\[
-0.0245 \pm \frac{0.000196}{\sqrt{9,999}} \cdot \sqrt{5,449,975} = -0.0245 \pm 0.0046 = [-0.0291, -0.0199].
\]

Note, as mentioned earlier, the half-width of this confidence interval is \(h = 0.0046\).

The above computation of the confidence interval is a bit nasty! Because \(n = 10,000\) is very large, the term under the square root sign is large and messy to obtain. When our ‘data’ come
from a computer simulation experiment, \( n \) is frequently very large. In these cases, there is an approximate formula which is much easier to use. In particular, consider the term under the square root. Its numerator is

\[
n(b + c) - (b - c)^2.
\]

For the simulation experiment above, this term becomes

\[
n(b + c) - (b - c)^2 = 10000(551) - (245)^2 = 5,449,975.
\]

The point of this argument is that the term \((b - c)^2\) has a negligible effect on the answer; with it, we obtain 5,449,975; without it, we would obtain 5,510,000. The former is only 1.1% smaller than the latter. If we exclude the \((b - c)^2\) term, then square root term in Formula 16.8 reduces to:

\[
\sqrt{\frac{n}{n-1}}(b + c).
\]

For \( n = 10,000 \), the ratio \( n/(n - 1) \) is almost one. If we ignore it, Formula 16.8 reduces to the much simpler formula:

\[
\left(\frac{b - c}{n}\right) \pm \left(\frac{z^*}{\sqrt{n}}\right)\sqrt{b + c},
\]

(16.9)

If I use this new formula for our earlier data, which had:

\[
b = 153; c = 398; n = 10,000,
\]

we get

\[
-0.0245 \pm 0.000196\sqrt{153 + 398} = -0.0245 \pm 0.0046;
\]

the same answer we obtained from Formula 16.8.

Let’s apply the above inference methods to the Larry Bird data in Table 15.16 in Chapter 15. I will reproduce his data below:

<table>
<thead>
<tr>
<th>First Shot</th>
<th>Second Shot</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>251</td>
<td>34</td>
</tr>
<tr>
<td>Miss</td>
<td>48</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>299</td>
<td>39</td>
</tr>
</tbody>
</table>

We see that

\[
b = 34; c = 48; b + c = m = 82; b - c = -14; \text{ and } n = 338.
\]

For the test of \( H_0 : P(A) = P(B) \), I choose the alternative \( \neq \). and go to the website:

http://stattrek.com/Tables/Binomial.aspx

I enter: 0.5 for the Probability of success; \( m = 82 \) for the Number of trials; and \( b = 34 \) for the Number of successes. I click on Calculate and obtain:

\[
P(X \geq 34) = 0.9515 \text{ and } P(X \leq 34) = 0.0753.
\]
Thus, the P-value for $>$ is 0.9515; the P-value for $<$ is 0.0753; and the P-value for my alternative, $\neq$ is $2(0.0753) = 0.1506$. There is evidence that Bird’s success probability on his second shot was larger than his success probability on his first shot, but the evidence is not convincing.

The approximate 95% confidence interval for $P(A) - P(B)$ is:

$$(-14/338) \pm (1.96/338)\sqrt{\frac{338(82) - (-14)^2}{337}} = -0.041 \pm 0.052 = [-0.093, 0.011].$$

This is a very wide interval; 5.2 percentage points—its half-width—is a substantial amount in free throw shooting. The point estimate, $-0.041$, suggests that Bird might have been much better on his second shots, but the data are inconclusive.

By the way, if I use the approximate (simpler) Formula 16.9, I obtain:

$$-0.041 \pm (1.96/338)\sqrt{34 + 48} = -0.041 \pm 0.053.$$

The approximation, while excellent, is not quite as good as before because $n = 338$ instead of 10,000. As a practical matter, however, to me a half-width of 0.053 has the same meaning as a half-width of 0.052.

16.5 Paired Data; Randomization-based Inference

In the largest sense, every example in this chapter has paired data. In one study, each person’s sex is paired with the person’s response on lenses. In another study, Casey’s barking behavior is paired with the foraging of squirrels. I would call neither of these, however, paired data. For me, in this chapter, I reserve the term paired data to situations in which I want to compare $P(A)$ and $P(B)$.

Two studies fit this criterion: our revisit of the Chapter 9 study of power and our analysis of free throw data.

There is another way to view paired data that is useful: the two examples in this chapter are cases of reusing units (or reusing subjects or trials). For example, in the study of power our subjects are assignments and we obtain two responses/features—i.e., we reuse—from each subject. Similarly, Larry Bird goes to the line to attempt a pair of free throws; the trial gives us his response on the first attempt and then we reuse it to obtain his response on the second attempt.

I chose to analyze Bird’s data with a population model. This means that the 338 trials in our data set are viewed as the realization of 338 i.i.d. trials. This extra assumption creates a mathematical structure in which the quantities $P(A)$ and $P(B)$ make sense; thus allowing us to derive various confidence interval estimates—for $P(A)$, $P(B)$ and $P(A) - P(B)$. Without a population model, I cannot use Bird’s data for inference because randomization is not possible—as you will soon see, randomizing would mean I could randomize the order of his shots, but clearly the first shot must be taken before the second shot.

My first example below is one that has been a favorite of teachers of Statistics throughout my career, but I have no idea whether such a study has ever been conducted! Imagine that we want to compare two therapies—call them cream 1 and cream 2—as a treatment for acne. As I hope these names suggest, the therapies would be applied to the skin as opposed to being taken orally or given in a shot. This will be important.
Arguably, a numerical response would be natural, but let’s assume that the response is a dichotomy, with possibilities improved—a success—and not improved—a failure. Suppose that we had 100 persons suffering from acne available for study. We could perform a CRD, using either randomization-based inference—Chapter 8—or population-based inference—Chapter 15. We could, however, do something else, which I will now describe.

We could reuse each subject. In particular, if Bert has acne, we could tell him to put cream 1 on the left side of his face and cream 2 on the right side of his face. After the period of treatment, we would obtain two responses from Bert. We might, for example, code the responses as:

- $A$ if cream 1 yields a success;
- $A^c$ if cream 1 yields a failure;
- $B$ if cream 2 yields a success; and
- $B^c$ if cream 2 yields a failure.

Here are two natural questions:

1. How did the researcher decide that cream 1 would be applied to the left side Bert’s face?
2. How about the other subjects in the study; Would they all apply cream 1 to the left sides of their faces?

Let’s consider the second question first. It would be bad science to have every subject put cream 1 on the left side and cream 2 on the right side. Well, perhaps I should say potentially bad science. I don’t really know whether side of the face influences the response. But if I performed the study in this way, all I can legitimately claim is that I have compared left-side-cream 1 to right-side-cream-2; in other words, the effect of the side of the face would be completely confounded with the effect of the type of cream.

There are two solutions to this side-of-the-face issue. The first is simple randomization. Suppose that there are 100 persons available for study. A separate randomization is performed for each of the 100 persons. For example, for Bert, side left or right is selected at random; cream 1 is applied to the selected side and cream 2 is applied to the remaining side. Thus, with 100 persons the overall randomization will involve 100 small randomizations.

Suppose that we do indeed have 100 subjects for study and we perform the 100 separate randomizations as described above. Let $X$ denote the number of subjects who will place cream 1 on the left sides of their faces. Clearly, the probability distribution of $X$ is Bin(100,0.5). We know from earlier work—or it can be easily verified with our binomial calculator website—that $P(X = 50)$ is small; only 0.0796. Thus, there is a 92% chance that one of the creams will be applied to more left sides than the other cream. Also, $P(40 \leq X \leq 60) = 0.9648$; thus, a discrepancy of more than 20 from side-to-side is unlikely, but not unimaginable. If the researcher feels strongly that side-of-face has a strong influence on the response, the fact that randomization could lead to one cream getting an additional 20 or more good sides is disturbing.

A solution to the above issue—namely, that randomization will likely lead to $X \neq 50$—is given by a cross-over design. In a cross-over design we select only one assignment, which is a desirable
simplification over the 100 needed above. The one assignment selects 50 subjects at random from the total of 100. Each of the 50 selected subjects puts cream 1 on the left side of the face and each of the remaining 50 subjects puts cream 1 on the right side of the face.

An obvious question is: Why not always use a cross-over design? The cross-over design requires a more complicated analysis and we don’t have time to present it in this course. (More accurately, I have made an executive decision not to present it.)

I conjecture that the above acne example has remained popular with teachers of Statistics because it is such a natural example of a medical issue that can be treated two different ways simultaneously. The next step in the hierarchy would be a medical issue which can be addressed with only one therapy at a time, but is recurrent. An obvious example—one familiar to readers of these notes—would be a study of headaches. Artificial Headache Study-2 (HS-2) was introduced in Example 8.4 on page 168 and its data are in Table 8.4. For convenience, I reproduce its data below with the names of the drugs changed to 1 and 2:

<table>
<thead>
<tr>
<th>Drug</th>
<th>Pain relieved?</th>
<th>Row Prop.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

The exact P-value for the alternative $\neq$ for Fisher’s test is 0.1612—details not given.

We could modify this headache study to enable subject reuse. In particular, we would need two headaches per subject, with one headache treated by drug 1 and the other with drug 2. The treatment assigned to the first headache would be determined by randomization.

I will create two distinct artificial data sets to investigate the issue of whether or not sample reuse is a good strategy. I need to be careful. I want to compare each new data set to the data in HS-2. Recall that HS-2 required 100 subjects—50 assigned to each treatment—to obtain the total of 100 observations in the data table above. One hundred subjects reused would give us data on 200 headaches, which seems to me to be giving an unfair advantage to subject reuse. Therefore, each of my two data sets for subject reuse has 50 subjects, giving a total of 100 headaches.

Next, I want both of my two new data sets to be comparable to the data from HS-2. Here is what I mean. With subject reuse, my $2 \times 2$ data table will look like the following.

<table>
<thead>
<tr>
<th>Drug 2</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Success</td>
<td>Failure</td>
<td></td>
</tr>
<tr>
<td>Success (A)</td>
<td>$a$</td>
<td>$b$</td>
<td>$n_1$</td>
</tr>
<tr>
<td>Failure (A')</td>
<td>$c$</td>
<td>$d$</td>
<td>$n_2$</td>
</tr>
<tr>
<td>Total</td>
<td>$m_1$</td>
<td>$m_2$</td>
<td>50</td>
</tr>
</tbody>
</table>

In HS-2, drug 1 gives 58% successes and drug 2 gives 42% successes. To be fair (comparable) I must have these same numbers for both of my subject reuse data sets. Thus, their data tables will look like:
Table 16.13: The two extreme subject reuse data sets consistent with HS-2 data. For the alternative $\neq$, the P-value for $b = 8$ is 0.0078 and the P-value for $b = 29$ is 0.3222. For comparison, the P-value for the HS-2 data with the same alternative is 0.1612.

<table>
<thead>
<tr>
<th></th>
<th>Smallest Possible $b$</th>
<th>Largest Possible $b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Drug 2</td>
<td>Drug 2</td>
</tr>
<tr>
<td></td>
<td>Success ($B$)</td>
<td>Failure($B^c$)</td>
</tr>
<tr>
<td>Succ. ($A$)</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>Fail. ($A^c$)</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>29</td>
</tr>
</tbody>
</table>

Recall that in these tables, $a$ counts the number of (reused) subjects who would achieve a success with both drugs; $d$ counts the number of (reused) subjects who would achieve a failure with both drugs; $b$ counts the number of (reused) subjects for whom—more picturesquely—drug 1 defeated drug 2; and $c$ counts the number of (reused) subjects for whom drug 1 lost to drug 2.

You may verify—or you may simply trust me—that in this last table, $b$ can take on any integer value from 8 to 29. I look at the extremes of these possibilities in Table 16.13. Let’s take a few minutes to examine the information in this table.

For $b = 8$, the data are highly statistically significant, with an exact P-value of 0.0078. For $b = 29$, the exact P-value is very large, 0.3222. First, although I won’t prove this, if I created similar tables for $b = 9, 10, 11, \ldots, 28$, we would find that the P-value for the alternative $\neq$—as well as the alternative $>$—would increase with the value of $b$. Indeed, we would find that for $b = 16$ the P-value is 0.1516 and for $b = 17$ the P-value is 0.1686. The table with $b = 17$ is:

<table>
<thead>
<tr>
<th></th>
<th>Success ($B$)</th>
<th>Failure($B^c$)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Succ. ($A$)</td>
<td>12</td>
<td>17</td>
<td>29</td>
</tr>
<tr>
<td>Fail. ($A^c$)</td>
<td>9</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>29</td>
<td>50</td>
</tr>
</tbody>
</table>

In this table, slightly fewer than one-half of the subjects (actual count, 24 of 50) respond the same to each drug and slightly more than one-half of the subjects respond differently to the drugs. As a physician, I can’t believe that the results would be dissimilar for so many subjects; thus, I would anticipate obtaining $b$ that is smaller than 17 and, hence, opt for subject reuse because it would give me a smaller P-value; i.e., it would be more sensitive. Of course, I am not a physician; thus, the validity of my opinion is quite questionable. Someday if you are the researcher in a study like this, you will need to decide whether to have subject reuse.
Above I have talked about sample reuse for hypothetical studies of acne and headaches. If one opts for subject reuse, then one can perform population-based inference or the randomization-based inference of Part I of these notes; from either perspective, one uses McNemar’s test and obtains the same P-value for any fixed alternative.

16.5.1 Maslow’s Hammer Revisited

Above I talked about a hierarchy: for acne, a subject can receive two treatments simultaneously; for headaches a subject can receive two treatments serially. Intuition often suggests—and experience has often verified—that subject reuse can lead to a more efficient study. It is perhaps then simply human nature (Maslow’s Hammer) that researchers seek other venues for subject reuse. In this subsection I will discuss one such situation, with lots of cautionary words for you.

Suppose that you have 200 patients with colon cancer and two therapies that you want to compare. The above ideas for acne and headaches clearly won’t work. But here is an idea. Before assigning subjects to treatments, record the values of several variables on each individual and summarize these values with a single number. Let’s assume that we believe that the larger the number, the better the subject’s prognosis for a favorable response, be it a dichotomy or a number. In this scenario it is valid to do the following:

Use the values of the 200 numbers to form 100 pairs of subjects in the following way. The subjects with the two largest numbers create a pair. Of the remaining 198 numbers, the subjects with the two largest numbers create a pair. And so on.

Once the 100 pairs are formed, within each pair select a subject at random to be given the first therapy; the other member of the pair will receive the second therapy.

Here is the really important point. If you proceed as above, then it is valid to perform randomization-based inference on the 100 pairs. I recommend against performing population-based inference in this case, but I don’t have time to give my whole argument, except to note the following. Suppose that Bert and Walt are among my 200 patients and that they are paired as above. If I looked at the entire population of hundreds of thousands of people (I am guessing at the population size), then I would be amazed if Bert and Walt were paired in the population. Thus, it is not clear how sample pairs relate to population pairs, so I won’t do it!

By the way, as you will see in Chapter 20, randomization is key. If one forms pairs in an observational study, the results are a disaster!

16.6 Summary

This chapter is concerned with the situation in which each unit (trial, subject) yields two dichotomous responses. We begin with units that are subjects; this situation, as before in these notes, leads us to define a finite population. We begin with the table of population counts, given in Table 16.1. If we divide each population count by the population size, \( N \), we obtain the table of population proportions or probabilities, given in Table 16.3.
In practice, these tables—counts and probabilities—are known only to Nature. When units are trials, there is no notion of population counts, making the table of probabilities the starting point. Again, in practice, the table of probabilities is known only to Nature.

The table of probabilities allows us to define the very useful and interesting concept of conditional probabilities, of which there are eight. Thus, in addition to the eight probabilities in the table of probabilities, we have eight conditional probabilities; 16 is a lot of probabilities! But, no worries: we learn that there are really only three non-redundant (conditional or not) probabilities. There are many valid sets of three non-redundant probabilities and any set of them will yield the remaining 13 probabilities.

I introduce you to a very important use of the above ideas: screening tests for a disease. Also, you learn the important Bayes’ rule (or formula), which allows us to reverse the direction of conditioning.

The next issue is: How to obtain data in order to perform inference on the various probabilities of interest. Three possibilities are considered; listed and described on page 398. Type 2 sampling— independent random samples from the rows—is mathematically equivalent to Type 3 sampling— independent random samples from the columns.

Of the 18 possible probabilities, most analyses focus on comparing the members of one or more of the following pairs: \( P(A) \) and \( P(B) \); \( p_1 \) and \( p_2 \); and \( p_1^* \) and \( p_2^* \). Note that when I write \( p_1 \) and \( p_2 \), I am reverting to Chapter 15 notation; in Chapter 16 notation, \( p_1 = P(B\mid A) \) and \( p_2 = P(B\mid A^c) \). I prefer the Chapter 15 notation because it’s not so messy! Also, when I write \( p_1^* \) and \( p_2^* \), this is Chapter 15 notation in which the populations are in the columns and the success is in the first row. Again, \( p_1^* \) and \( p_2^* \) could be expressed in the messier conditional probability notation of Chapter 16.

The main result is that

1. For Type 1 sampling, all six of these parameters can be estimated.

2. For Type 2 sampling, only \( p_1 \) and \( p_2 \) can be estimated; the other four cannot be estimated.

3. For Type 3 sampling, only \( p_1^* \) and \( p_2^* \) can be estimated; the other four cannot be estimated.

I give a numerical example with Type 2 sampling that illustrates the second (and the mathematically equivalent third) of these results.

I then introduce you to a class of medical studies—the relationship between a risk factor and a bad outcome—for which we need to estimate \( p_1 \) and \( p_2 \), but neither Type 1 nor Type 2 sampling is realistic. Provided that \( p_1 \) and \( p_2 \) are relatively small, this difficulty can be overcome by focusing on the odds ratio rather than the relative risk, \( p_1 / p_2 \). We can then use Type 3 sampling to estimate the odds ratio, which is the same number for columns as it is for rows.

I give two examples in which the researcher is interested in comparing \( P(A) \) and \( P(B) \). It is easy to find the exact P-value for the test of the null hypothesis that these two probabilities are equal. The test is called McNemar’s test and is a special case of the test we learned in Chapter 12. There is also an approximate confidence interval estimate of \( P(A) - P(B) \), given in Formula 16.8, that is based on a Normal curve approximation.

The chapter ends with a lengthy discussion of applications to medical studies. We consider studies with subject reuse and randomization. It makes sense that subject reuse should lead to a
more sensitive analysis, but in any given situation the researcher must decide which method to use; there are no guarantees in this area!

16.7 Practice Problems

1. Suppose that we are given the following table of population counts:

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Bᶜ</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>800</td>
<td>200</td>
<td>1,000</td>
</tr>
<tr>
<td>Aᶜ</td>
<td>1,200</td>
<td>2,800</td>
<td>4,000</td>
</tr>
<tr>
<td>Total</td>
<td>2,000</td>
<td>3,000</td>
<td>5,000</td>
</tr>
</tbody>
</table>

   (a) Calculate the table of population probabilities.
   (b) Calculate the eight conditional probabilities.

2. Suppose that we are given the following table of population counts:

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Bᶜ</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>400</td>
<td>600</td>
<td>1,000</td>
</tr>
<tr>
<td>Aᶜ</td>
<td>1,600</td>
<td>2,400</td>
<td>4,000</td>
</tr>
<tr>
<td>Total</td>
<td>2,000</td>
<td>3,000</td>
<td>5,000</td>
</tr>
</tbody>
</table>

   (a) Calculate the table of population probabilities.
   (b) Calculate the eight conditional probabilities.
   (c) Notice that every conditional probability is equal to the corresponding unconditional probability. Whenever this happens, we say that the two responses are statistically independent. With independence, the multiplication rule for conditional probabilities:

\[ P(AB) = P(A)P(B|A), \text{ becomes } P(AB) = P(A)P(B), \]

which corresponds to our definition of independence in Chapter 10. In words, the probability associated with cell AB is the product of its row and column marginal probabilities. It now follows that the two responses are statistically independent if, and only if, every one of the four cell probabilities is equal to the product of its row and column marginal probabilities.

Of course, it’s no fun to check this multiplication for every cell. Fortunately, it can be shown that this multiplicative relationship holds for either: all four cells or none of the cells. Thus, we need to check only one cell.
3. Below is a table of probabilities.

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>0.195</td>
<td>0.455</td>
<td>0.650</td>
</tr>
<tr>
<td>$A^c$</td>
<td>0.105</td>
<td>0.245</td>
<td>0.350</td>
</tr>
<tr>
<td>Total</td>
<td>0.300</td>
<td>0.700</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Explain why you don’t need to do any calculations to obtain the eight conditional probabilities.

4. On page 392, I stated that if we know:

   A conditional probability for each row plus one of the row marginal probabilities, then we could obtain all eight of the probabilities in the table of probabilities. I will not prove this fact in all of its algebraic glory; instead, I will show you an example of how to do it.

   For example, given $P(B|A) = 0.375$, $P(B|A^c) = 0.500$ and $P(A) = 0.800$, determine the eight probabilities in Table 16.3.

   By the way, I also stated that if we know:

   A conditional probability for each column plus one of the column marginal probabilities, then we could obtain all eight of the probabilities in the table of probabilities. I won’t show you an example of this because it is just like the one I do show you, but with the rows and columns interchanged.

5. Below is the table of population counts for a disease and its screening test. (Recall that $A$ means the disease is present and $B$ means the screening test is positive.)

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>254</td>
<td>35</td>
<td>289</td>
</tr>
<tr>
<td>$A^c$</td>
<td>199</td>
<td>3126</td>
<td>3325</td>
</tr>
<tr>
<td>Total</td>
<td>453</td>
<td>3161</td>
<td>3614</td>
</tr>
</tbody>
</table>

   (a) What proportion of the population would test positive?
   (b) What proportion of the population is disease free?
   (c) What proportion of the population is free of the disease and would test negative?
   (d) What proportion of the population has the disease and would test positive?
   (e) Of those who would test negative, what proportion has the disease?
   (f) Of those who are free of the disease, what proportion would test positive?
   (g) What proportion of the population would receive a correct screening test result?
(h) Of those who would receive an incorrect screening test result, what proportion would receive a false positive?

(i) What proportion of the population does not have the disease or would test negative?

6. My dog Casey would visit my neighbor Sally while she was shooting free throws. I could see Sally shoot, but I could not see the outcome of her shot. Because Sally was a professional poker player, she did not have a *tell*; i.e., as best I could discern, her reaction to a made shot was identical to her reaction to a missed shot. In other words, Sally and Casey could see everything; I could only see when shots were attempted. According to Sally, her free throws were Bernoulli trials with probability of success (made shot) \( p = 0.80 \). Also according to Sally, immediately after she made a shot, Casey would bark 70\% of the time; immediately after she missed a shot, Casey would bark 40\% of the time (Sally liked to feed squirrels). Casey would neither confirm nor refute Sally’s numbers.

Use the above information to answer the following questions.

(a) Sally is preparing to shoot; what is the probability that Casey is about to bark?

(b) Sally has shot and Casey has barked; what is the probability Sally made the shot?

(c) Sally has shot and Casey is silent; what is the probability Sally made the shot?

7. Refer to the medical studies introduced in Section 16.3. The population counts are given by the following table, in thousands.

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>9</td>
<td>291</td>
<td>300</td>
</tr>
<tr>
<td>( A^c )</td>
<td>7</td>
<td>693</td>
<td>700</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>984</td>
<td>1,000</td>
</tr>
</tbody>
</table>

Calculate the values of \( p_1, p_2 \), the relative risk and the odds ratio for this population.

8. Refer to the previous problem. I used Minitab to generate data from this population using a case-control study (Type 3 sampling) and obtained the following data.

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>99</td>
<td>93</td>
<td>192</td>
</tr>
<tr>
<td>( A^c )</td>
<td>101</td>
<td>207</td>
<td>308</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>300</td>
<td>500</td>
</tr>
</tbody>
</table>

(a) Calculate the odds ratio for these data. Viewing this number as the point estimate of the population odds ratio, comment.

(b) Obtain the approximate 95\% confidence interval estimate of the population odds ratio. Is the interval estimate correct?
9. The data for Rick Roby shooting free throws is below:

<table>
<thead>
<tr>
<th>First Shot</th>
<th>Second Shot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit</td>
</tr>
<tr>
<td>Hit</td>
<td>54</td>
</tr>
<tr>
<td>Miss</td>
<td>49</td>
</tr>
<tr>
<td>Total</td>
<td>103</td>
</tr>
</tbody>
</table>

(a) Find the exact P-values for each of the three possible alternatives for the test of the null hypothesis that \( P(A) = P(B) \).

(b) Calculate the approximate 95% confidence interval estimate of \( P(A) - P(B) \).

16.8 Solutions to Practice Problems

1. (a) I divide each count by \( N = 5,000 \) and obtain the table of probabilities below:

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>0.16</td>
<td>0.04</td>
<td>0.20</td>
</tr>
<tr>
<td>( A^c )</td>
<td>0.24</td>
<td>0.56</td>
<td>0.80</td>
</tr>
<tr>
<td>Total</td>
<td>0.40</td>
<td>0.60</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(b) The table of conditional probabilities of \( B \)'s given \( A \)'s is below:

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>( P(B</td>
<td>A) = 0.80 )</td>
<td>( P(B^c</td>
</tr>
<tr>
<td>( A^c )</td>
<td>( P(B</td>
<td>A^c) = 0.30 )</td>
<td>( P(B^c</td>
</tr>
</tbody>
</table>

The table of conditional probabilities of \( A \)'s given \( B \)'s is below. For example, \( P(A|B^c) = 0.07 \).

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>( P(A</td>
<td>B) = 0.40 )</td>
<td>( P(A</td>
</tr>
<tr>
<td>( A^c )</td>
<td>( P(A^c</td>
<td>B) = 0.60 )</td>
<td>( P(A^c</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

2. (a) I divide each count by \( N = 5,000 \) and obtain the table of probabilities below:

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>0.08</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>( A^c )</td>
<td>0.32</td>
<td>0.48</td>
<td>0.80</td>
</tr>
<tr>
<td>Total</td>
<td>0.40</td>
<td>0.60</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(b) The table of conditional probabilities of \( B \)'s given \( A \)'s is below.
The table of conditional probabilities of $A$’s given $B$’s is below.

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$P(B</td>
<td>A) = 0.40$</td>
<td>$P(B^c</td>
</tr>
<tr>
<td>$A^c$</td>
<td>$P(B</td>
<td>A^c) = 0.40$</td>
<td>$P(B^c</td>
</tr>
<tr>
<td></td>
<td>$P(B) = 0.40$</td>
<td>$P(B^c) = 0.60$</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The table of conditional probabilities of $A$’s given $B$’s is below.

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$P(A</td>
<td>B) = 0.20$</td>
<td>$P(A</td>
</tr>
<tr>
<td>$A^c$</td>
<td>$P(A^c</td>
<td>B) = 0.80$</td>
<td>$P(A^c</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

3. We can see that $P(AB) = 0.195$ is equal to $P(A)P(B) = 0.650(0.300) = 0.195$. Thus, the two responses are independent (refer to the previous practice problem). As a result the marginal probabilities are equal to the conditional probabilities. For example

$$0.30 = P(B) = P(B|A) = P(B|A^c).$$

4. We proceed as follows. Given that $P(A) = 0.800$, we know that $P(A^c) = 0.200$. We put these two numbers into our table of probabilities:

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>0.800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A^c$</td>
<td>0.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next, we use my second statement of the multiplication rule for conditional probabilities, Equation [16.4] twice. First,

$$P(AB) = P(A)P(B|A) = 0.800(0.375) = 0.300.$$ Second, we use it with a slight change of names:

$$P(A^cB) = P(A^c)P(B|A^c) = 0.200(0.500) = 0.100.$$ We place these two newly acquired probabilities into our table, giving:

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>0.300</td>
<td>0.800</td>
<td></td>
</tr>
<tr>
<td>$A^c$</td>
<td>0.100</td>
<td>0.200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, after a flurry of additions and subtractions, we obtain:

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>0.300</td>
<td>0.500</td>
<td>0.800</td>
</tr>
<tr>
<td>$A^c$</td>
<td>0.100</td>
<td>0.100</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.600</td>
<td>1.000</td>
</tr>
</tbody>
</table>
5. (a) $453/3614 = 0.125$; (b) $3325/3614 = 0.920$; (c) $3126/3614 = 0.865$; (d) $254/3614 = 0.070$; (e) $35/3161 = 0.011$; (f) $199/3325 = 0.060$; (g) $(254 + 3126)/3614 = 0.935$; (h) $199/(35 + 199) = 0.850$; (i) $(3614 - 254)/3614 = 0.930$.

6. This, of course, is a dreaded story problem. The primary challenge is to write the given information in the language of this chapter. First, we identify the trial: Sally attempts a free throw. Second, we identify the two dichotomous responses and give them labels: one response is the outcome of the shot and the other response is Casey’s behavior. I will define $B$ for Casey barking and $A$ for Sally making her shot. This gives us the following table of probabilities with unknown probabilities missing:

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>$A^c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.80</td>
<td>0.20</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Next, we write one piece of the given information in symbols: $P(A) = 0.80$, which implies that $P(A^c) = 0.20$. We put this information in our table:

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>0.80</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>$A^c$</td>
<td></td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.64</td>
<td>0.36</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Next, we use the multiplication rule for conditional probabilities twice:

$$P(AB) = P(A)P(B|A) = 0.80(0.70) = 0.56$$ and

$$P(A^cB) = P(A^c)P(B|A^c) = 0.20(0.40) = 0.08.$$ 

We put this information in our table and do lots of adding and subtracting; the end result is:

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$B^c$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>0.56</td>
<td>0.24</td>
<td>0.80</td>
</tr>
<tr>
<td>$A^c$</td>
<td>0.08</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>Total</td>
<td>0.64</td>
<td>0.36</td>
<td>1.00</td>
</tr>
</tbody>
</table>

We are now ready to answer questions!

(a) The question asks for $P(B)$; from the table, $P(B) = 0.64$.

(b) The question asks for $P(A|B)$; from the definition of conditional probability,

$$P(A|B) = P(AB)/P(B),$$ which equals $0.56/0.64 = 0.875$. 

421
(c) The question asks for $P(A|B^c)$; from the definition of conditional probability,

$$P(A|B^c) = P(AB^c)/P(B^c),$$

which equals $0.24/0.36 = 0.667$.

7. We have $p_1 = 9/300 = 0.03$ and $p_2 = 7/700 = 0.01$. The relative risk is $p_1/p_2 = 0.03/0.01 = 3$. The odds ratio is

$$\theta = \frac{9(693)}{291(7)} = 3.062.$$

8. (a) The odds ratio of these data is

$$\hat{\theta} = \frac{99(207)}{93(101)} = 2.182.$$

The point estimate is considerably smaller than the population value, 3.062.

(b) First, I compute

$$\hat{\lambda} = \ln(\hat{\theta}) = \ln(2.182) = 0.7802.$$

The 95% confidence interval estimate of $\lambda$ is

$$0.7802 \pm 1.96\sqrt{\frac{1}{99} + \frac{1}{93} + \frac{1}{101} + \frac{1}{207}} = 0.7802 \pm 1.96(0.1886) = 0.7802 \pm 0.3697 = [0.4105, 1.1499].$$

This gives us the following 95% confidence interval estimate of $\theta$:

$$e^{0.4105} \leq \theta \leq e^{1.1499}$$

or $1.508 \leq \theta \leq 3.158$.

This interval is correct because it (barely) includes $\theta = 3.062$.

9. (a) Go to the website [http://stattrek.com/Tables/Binomial.aspx](http://stattrek.com/Tables/Binomial.aspx) to obtain the exact P-values for McNemar’s test.

After accessing the site, enter 0.5 as the Probability of success; enter $m = 37 + 49 = 86$ for the Number of trials; and enter $b = 37$ for the Number of successes. Click on calculate and you will obtain five probabilities. The two that are relevant are:

- $P(X \leq 37) = 0.1177$
- $P(X \geq 37) = 0.9197$.

Thus, the P-value for < is 0.1177; the P-value for > is 0.9197; and the P-value for $\neq$ is $2(0.1177) = 0.2354$.

(b) For the confidence interval estimate, note that $(b - c) = -12$, $b + c = 86$ and $n = 171$. Thus, the interval is

$$\left(\frac{-12}{171}\right) \pm \left(1.96/171\right)\sqrt{\frac{171(86) - (-12)^2}{170}} = -0.0702 \pm 0.01146(9.255) = -0.0702 \pm 0.1061 = [-0.1763, 0.0359].$$

In the data, Roby was a considerably better shooter on his second shot, but the confidence interval estimate is inconclusive.
16.9 Homework

1. This is a problem about a very good screening test for a very rare disease. You are given the following probabilities:

\[ P(A) = 0.001, \ P(B|A) = 0.999 \ \text{and} \ \ P(B|A^c) = 0.01. \]

Calculate \( P(A^c|B) \). Comment.

Hint: Rather than work with very small probabilities, it might be easier to work with population counts. To this end, let the population size \( N \) be one million.

2. This problem is about relative risks and odds ratios. Below is a table of hypothetical population counts, in thousands.

<table>
<thead>
<tr>
<th>Group</th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>12</td>
<td>188</td>
<td>200</td>
</tr>
<tr>
<td>( A^c )</td>
<td>12</td>
<td>788</td>
<td>800</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>976</td>
<td>1000</td>
</tr>
</tbody>
</table>

A case-control study with 800 subjects from this population yielded the data below.

<table>
<thead>
<tr>
<th>Group</th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>207</td>
<td>90</td>
<td>297</td>
</tr>
<tr>
<td>( A^c )</td>
<td>193</td>
<td>310</td>
<td>503</td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td>400</td>
<td>800</td>
</tr>
</tbody>
</table>

(a) Calculate the relative risk and odds ratio for the population.

(b) Calculate the point estimate of the population odds ratio.

(c) Obtain the 95% CI for the population odds ratio.

3. A former student of mine, Jackie, planned to study her dog, Basia, with a total of 50 trials. Jackie wanted to study Basia’s ability to catch a kernel of popped corn that has been tossed towards her. (I am guessing that Basia is a female.) Years of experience had convinced Jackie that Basia was very skilled at catching popcorn that was tossed directly at her. For her study, Jackie chose the following two treatments. For the first [second] treatment, Jackie would toss the popcorn approximately two feet to Basia’s right [left]. A trial is labeled a success if, and only if, Basia catches the kernel before it hits the ground.

Jackie, of course, could perform a CRD and analyze it using the methods of Chapter 15—if she was willing to assume Bernoulli trials—or Chapter 8—if not. Here is another idea. We can take the 50 trials and form 25 pairs. Trials 1 and 2 form the first pair; trials 3 and 4 form the second pair; and so on. Jackie chose to perform this randomized pairs design; similar to the acne and headache studies, Jackie performed a separate randomization for each of her 25 pairs of trials. This is mathematically valid because Jackie it is a form of reusing units.
Jackie obtained the following results: Basia obtained a total of 16 successes on the first treatment; seven pairs of trials yielded two successes; and four pairs of trials yielded two failures.

(a) Use the information above to complete the following data table:

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1:</th>
<th>Treatment 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Success</td>
<td>Success</td>
<td>25</td>
</tr>
<tr>
<td>Success</td>
<td></td>
<td>Failure</td>
<td></td>
</tr>
<tr>
<td>Failure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>25</td>
</tr>
</tbody>
</table>

(b) Obtain the exact P-value for each of the three possible alternatives.

(c) Now pretend that Jackie had performed a CRD and obtained exactly the same data. Obtain the exact P-value for each of the three possible alternatives. Compare these answers to your answers in (b).

4. Refer to the previous problem. Leigh performed 50 pairs of trials on her dog Attica. A trial consisted of Leigh standing near a window inside her home while Attica was reposed on the floor. For treatment 1, Leigh would yell, “Squirrel, Attica!” For treatment 2, Leigh would calmly remark, “Hey Attica, squirrel.” Attica’s response was classified into one of two categories—a success if she got excited and a failure if she did not move. There was a total of 37 successes on the first treatment and a total of only 16 successes on the second treatment. For only five pairs of trials did Attica give two successes.

(a) Present these data in a $2 \times 2$ table.

(b) Find the exact P-value for the alternative $>$. 

(c) Pretend that Leigh had performed a CRD and obtained exactly the same data. Obtain the exact P-value for the alternative $>$. 

424