*BIOSTATISTICS\_Daniel\_C03*

***CHAPTER OVERVIEW***

*Probability lays the foundation for statistical inference.This chapter provides a brief overview of the probability concepts necessary for the understanding of topics covered in the chapters that follow. It also provides a context for understanding the probability distributions used in statistical inference, and introduces the student to several measures commonly found in the medical literature (e.g., the sensitivity and specificity of a test).*

*La probabilidad sienta las bases de la inferencia estadística. Este capítulo ofrece una breve descripción general de los conceptos de probabilidad necesarios para comprender los temas que se abordan en los capítulos siguientes. También proporciona un contexto para comprender las distribuciones de probabilidad utilizadas en la inferencia estadística y presenta al estudiante diversas medidas comunes en la literatura médica (p. ej., la sensibilidad y la especificidad de una prueba).*

***TOPICS***

*3.1 INTRODUCTION  
   
3.2 TWO VIEWS OF PROBABILITY: OBJECTIVE AND SUBJECTIVE  
   
3.3 ELEMENTARY PROPERTIES OF PROBABILITY   
3.4 CALCULATING THE PROBABILITY OF AN EVENT  
   
3.5 BAYES’ THEOREM, SCREENINGTESTS, SENSITIVITY, SPECIFICITY, AND PREDICTIVE VALUE POSITIVE AND NEGATIVE*

*3.6 SUMMARY LEARNING OUTCOMES*

***LEARNING OUTCOMES***

*After studying this chapter, the student will*

1. *understand classical, relative frequency, and subjective probability.*
2. *understand the properties of probability and selected probability rules.*
3. *be able to calculate the probability of an event.*
4. *be able to apply Bayes’ theorem when calculating screening test results.*

***3.1 INTRODUCTION***

*The theory of probability provides the foundation for statistical inference. However, this theory, which is a branch of mathematics, is not the main concern of this book, and, consequently, only its fundamental concepts are discussed here. Students who desire to pursue this subject should refer to the many books on probability available in most college and university libraries. The books by Gut (1), Isaac (2), and Larson (3) are recommended. The objectives of this chapter are to help students gain some mathematical ability in the area of probability and to assist them in developing an understanding of the more impor- tant concepts. Progress along these lines will contribute immensely to their success in under- standing the statistical inference procedures presented later in this book.*

*The concept of probability is not foreign to health workers and is frequently encountered in everyday communication. For example, we may hear a physician say that a patient has a 50–50 chance of surviving a certain operation. Another physician may say that she is 95 percent certain that a patient has a particular disease. A public health nurse may say that nine times out of ten a certain client will break an appointment. As these examples suggest, most people express probabilities in terms of percentages. In dealing with probabilities mathematically, it is more convenient to express probabilities as fractions. (Percentages result from multiplying the fractions by 100.) Thus, we meas- ure the probability of the occurrence of some event by a number between zero and one. The more likely the event, the closer the number is to one; and the more unlikely the event, the closer the number is to zero. An event that cannot occur has a probability of zero, and an event that is certain to occur has a probability of one.*

*Health sciences researchers continually ask themselves if the results of their efforts could have occurred by chance alone or if some other force was operating to produce the observed effects. For example, suppose six out of ten patients suffering from some disease are cured after receiving a certain treatment. Is such a cure rate likely to have occurred if the patients had not received the treatment, or is it evidence of a true curative effect on the part of the treatment? We shall see that questions such as these can be answered through the application of the concepts and laws of probability.*

***3.2 TWO VIEWS OF PROBABILITY: OBJECTIVE AND SUBJECTIVE***

*Until fairly recently, probability was thought of by statisticians and mathematicians only as an objective phenomenon derived from objective processes.*

*The concept of objective probability may be categorized further under the headings of (1) classical, or a priori, probability; and (2) the relative frequency, or a posteriori, con- cept of probability.*

***Classical Probability*** *The classical treatment of probability dates back to the 17th century and the work of two mathematicians, Pascal and Fermat. Much of this the- ory developed out of attempts to solve problems related to games of chance, such as those involving the rolling of dice. Examples from games of chance illustrate very well he principles involved in classical probability. For example, if a fair six-sided die is rolled, the probability that a 1 will be observed is equal to 1>6 and is the same for the other five faces. If a card is picked at random from a well-shuffled deck of ordinary playing cards, the probability of picking a heart is 13>52. Probabilities such as these are calculated by the processes of abstract reasoning. It is not necessary to roll a die or draw a card to compute these probabilities. In the rolling of the die, we say that each of the six sides is equally likely to be observed if there is no reason to favor any one of the six sides. Similarly, if there is no reason to favor the drawing of a particular card from a deck of cards, we say that each of the 52 cards is equally likely to be drawn. We may define probability in the classical sense as follows:*

***DEFINITION***

*If an event can occur in N mutually exclusive and equally likely ways, and if m of these possess a trait E, the probability of the occurrence of E is equal to m􏰉N.*

*If we read P(E) as “the probability of E,” we may express this definition as*

**

***Relative Frequency Probability*** *The relative frequency approach to prob- ability depends on the repeatability of some process and the ability to count the number of repetitions, as well as the number of times that some event of interest occurs. In this context we may define the probability of observing some characteristic, E, of an event as follows:*

***DEFINITION***

*If some process is repeated a large number of times, n, and if some resulting event with the characteristic E occurs m times, the relative frequency of occurrence of E, m>n, will be approximately equal to the probability of E.*

*To express this definition in compact form, we write*

**

*We must keep in mind, however, that, strictly speaking, m􏰉n is only an estimate of P1E2.*

***Subjective Probability*** *In the early 1950s, L. J. Savage (4) gave considerable impetus to what is called the “personalistic” or subjective concept of probability. This view holds that probability measures the confidence that a particular individual has in the truth of a particular proposition. This concept does not rely on the repeatability of any process. In fact, by applying this concept of probability, one may evaluate the probability of an event that can only happen once, for example, the probability that a cure for cancer will be discovered within the next 10 years.*

*Although the subjective view of probability has enjoyed increased attention over the years, it has not been fully accepted by statisticians who have traditional orientations.*

***Bayesian Methods*** *Bayesian methods are named in honor of the Reverend Thomas Bayes (1702–1761), an English clergyman who had an interest in mathematics. Bayesian methods are an example of subjective probability, since it takes into consider- ation the degree of belief that one has in the chance that an event will occur. While probabilities based on classical or relative frequency concepts are designed to allow for decisions to be made solely on the basis of collected data, Bayesian methods make use of what are known as prior probabilities and posterior probabilities.*

***DEFINITION***

*The prior probability of an event is a probability based on prior knowledge, prior experience, or results derived from prior data collection activity.*

***DEFINITION***

*The posterior probability of an event is a probability obtained by using new information to update or revise a prior probability.*

*As more data are gathered, the more is likely to be known about the “true” probability of the event under consideration. Although the idea of updating probabilities based on new information is in direct contrast to the philosophy behind frequency-of-occurrence proba- bility, Bayesian concepts are widely used. For example, Bayesian techniques have found recent application in the construction of e-mail spam filters. Typically, the application of Bayesian concepts makes use of a mathematical formula called Bayes’ theorem. In Section 3.5 we employ Bayes’ theorem in the evaluation of diagnostic screening test data.*

***3.3 ELEMENTARY PROPERTIES OF PROBABILITY***

*In 1933 the axiomatic approach to probability was formalized by the Russian mathe- matician A. N. Kolmogorov (5). The basis of this approach is embodied in three prop- erties from which a whole system of probability theory is constructed through the use of mathematical logic. The three properties are as follows.*

*1. Given some process (or experiment) with n mutually exclusive outcomes (called events), E1, E2, . . . , En, the probability of any event Ei is assigned a nonnegative number. That is,*

*P(Ei) >= 0 (3.3.1)*

*In other words, all events must have a probability greater than or equal to zero, a reasonable requirement in view of the difficulty of conceiving of negative probability. A key concept in the statement of this property is the concept of mutu- ally exclusive outcomes. Two events are said to be mutually exclusive if they can- not occur simultaneously.*

*2. The sum of the probabilities of the mutually exclusive outcomes is equal to 1.*

*P1E12 + P1E22 + Á + P1En2 = 1 (3.3.2)*

*This is the property of exhaustiveness and refers to the fact that the observer of a probabilistic process must allow for all possible events, and when all are taken together, their total probability is 1. The requirement that the events be mutually exclusive is specifying that the events E1, E2, . . . , En do not overlap; that is, no two of them can occur at the same time.*

*3. Consider any two mutually exclusive events, Ei and Ej. The probability of the occur- rence of either Ei or Ej is equal to the sum of their individual probabilities.*

*P1Ei + Ej2 = P1Ei2 + P1Ej2 (3.3.3)*

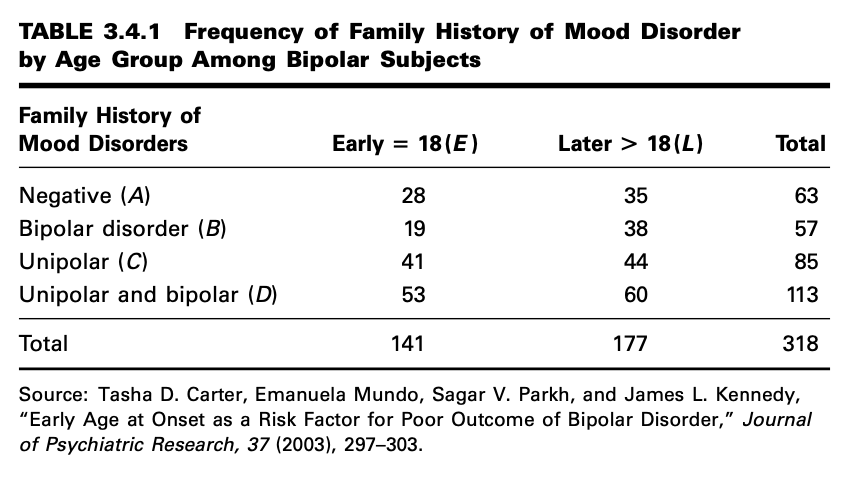
*Suppose the two events were not mutually exclusive; that is, suppose they could occur at the same time. In attempting to compute the probability of the occurrence of either Ei or Ej the problem of overlapping would be discovered, and the procedure could become quite complicated. This concept will be discussed further in the next section.*

***3.4 CALCULATING THE PROBABILITY OF AN EVENT***

*We now make use of the concepts and techniques of the previous sections in calculat- ing the probabilities of specific events. Additional ideas will be introduced as needed.*

***EXAMPLE 3.4.1***

*The primary aim of a study by Carter et al. (A-1) was to investigate the effect of the age at onset of bipolar disorder on the course of the illness. One of the variables investigated was family history of mood disorders. Table 3.4.1 shows the frequency of a family his- tory of mood disorders in the two groups of interest (Early age at onset defined to be*

**

*18 years or younger and Later age at onset defined to be later than 18 years). Suppose we pick a person at random from this sample. What is the probability that this person will be 18 years old or younger?*

*Solution:*

*For purposes of illustrating the calculation of probabilities, we consider this group of 318 subjects to be the largest group for which we have an interest. In other words, for this example, we consider the 318 subjects as a population. We assume that Early and Later are mutually exclusive cat- egories and that the likelihood of selecting any one person is equal to the likelihood of selecting any other person. We define the desired probability as the number of subjects with the characteristic of interest (Early) divided by the total number of subjects. We may write the result in probability notation as follows:*

*P1E2 = number of Early subjects /total number of subjects*

*= 141>318 = .4434 s*

***Conditional Probability*** *On occasion, the set of “all possible outcomes” may constitute a subset of the total group. In other words, the size of the group of interest may be reduced by conditions not applicable to the total group. When probabilities are calculated with a subset of the total group as the denominator, the result is a conditional probability.*

*The probability computed in Example 3.4.1, for example, may be thought of as an unconditional probability, since the size of the total group served as the denominator. No conditions were imposed to restrict the size of the denominator. We may also think of this probability as a marginal probability since one of the marginal totals was used as the numerator.*

*We may illustrate the concept of conditional probability by referring again to Table 3.4.1.*

***EXAMPLE 3.4.2***

*Suppose we pick a subject at random from the 318 subjects and find that he is 18 years or younger 1E2. What is the probability that this subject will be one who has no family history of mood disorders 1A2?*

*Solution:*

*The total number of subjects is no longer of interest, since, with the selec- tion of an Early subject, the Later subjects are eliminated. We may define the desired probability, then, as follows: What is the probability that a sub- ject has no family history of mood disorders 1A2, given that the selected subject is Early 1E2? This is a conditional probability and is written as P1A ƒ E2 in which the vertical line is read “given.” The 141 Early subjects become the denominator of this conditional probability, and 28, the num- ber of Early subjects with no family history of mood disorders, becomes the numerator. Our desired probability, then, is*

*P1A ƒ E2 = 28>141 = .1986 s*

***Joint Probability*** *Sometimes we want to find the probability that a subject picked at random from a group of subjects possesses two characteristics at the same time. Such a probability is referred to as a joint probability. We illustrate the calculation of a joint probability with the following example.*

***EXAMPLE 3.4.3***

*Let us refer again to Table 3.4.1. What is the probability that a person picked at random from the 318 subjects will be Early 1E2 and will be a person who has no family history of mood disorders 1A2?*

*Solution:*

*The probability we are seeking may be written in symbolic notation as P1E ̈ A2 in which the symbol ̈ is read either as “intersection” or “and.” The statement E ̈ A indicates the joint occurrence of conditions E and A. The number of subjects satisfying both of the desired conditions is found in Table 3.4.1 at the intersection of the column labeled E and the row labeled A and is seen to be 28. Since the selection will be made from the total set of subjects, the denominator is 318. Thus, we may write the joint probability as*

*P1E ̈ A2 = 28>318 = .0881 s*

***The Multiplication Rule*** *A probability may be computed from other prob- abilities. For example, a joint probability may be computed as the product of an appro- priate marginal probability and an appropriate conditional probability. This relationship is known as the multiplication rule of probability. We illustrate with the following example.*

*EXAMPLE 3.4.4*

*We wish to compute the joint probability of Early age at onset 1E2 and a negative family history of mood disorders 1A2 from knowledge of an appropriate marginal probability and an appropriate conditional probability.*

*Solution:*

*The probability we seek is P(E ̈ A). We have already computed a mar- ginal probability, P(E) = 141>318 = .4434, and a conditional probability, P1A ƒ E2 = 28>141 = .1986. It so happens that these are appropriate mar- ginal and conditional probabilities for computing the desired joint proba- bility. We may now compute P(E ̈ A) = P (E)P (A ƒ E) = 1.443421.19862 = .0881. This, we note, is, as expected, the same result we obtained earlier for P(E ̈ A). s*

*We may state the multiplication rule in general terms as follows: For any two events A and B,*

*P1A ̈ B2 = P1B2P1A ƒ B2, if P1B2 Z 0 (3.4.1)*

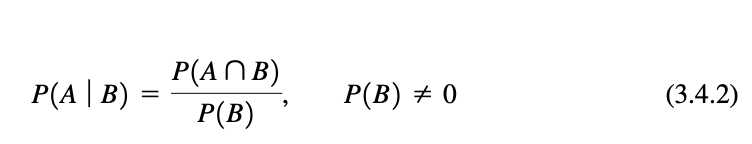
*For the same two events A and B, the multiplication rule may also be written as P1A ̈B2 = P1A2P1B ƒ A2,ifP1A2 Z 0.*

*We see that through algebraic manipulation the multiplication rule as stated in Equation 3.4.1 may be used to find any one of the three probabilities in its statement if the other two are known. We may, for example, find the conditional probability P1A ƒ B2 by dividing P1A ̈ B2 by P1B2. This relationship allows us to formally define conditional probability as follows.*

***DEFINITION***

*The conditional probability of A given B is equal to the probability of A o B divided by the probability of B, provided the probability of B is not zero.*

*That is,*

**

*We illustrate the use of the multiplication rule to compute a conditional probability with the following example.*

***EXAMPLE 3.4.5***

*We wish to use Equation 3.4.2 and the data in Table 3.4.1 to find the conditional probability, P1A ƒ E2.*

*Solution: According to Equation 3.4.2,*

*P1A|E2 = P1A ̈E2>P1E2*

*Earlier we found P1E ̈ A2 = P1A ̈ E2 = 28>318 = .0881. We have also determined that P1E2 = 141>318 = .4434. Using these results we are able to compute P1A ƒ E2 = .0881>.4434 = .1987, which, as expected, is the same result we obtained by using the frequencies directly from Table 3.4.1. (The slight discrepancy is due to rounding.)*

***The Addition Rule*** *The third property of probability given previously states that the probability of the occurrence of either one or the other of two mutually exclusive events is equal to the sum of their individual probabilities. Suppose, for example, that we pick a person at random from the 318 represented in Table 3.4.1. What is the probability that this person will be Early age at onset 1E2 or Later age at onset 1L2? We state this probability in symbols as P1E ́ L2, where the symbol*

*́ is read either as “union” or “or.” Since the two age conditions are mutually exclusive, P1E ̈ L2 = 1141>3182 + 1177>3182 = .4434 + .5566 = 1.*

*What if two events are not mutually exclusive? This case is covered by what is known as the addition rule, which may be stated as follows:*

***DEFINITION***

*Given two events A and B, the probability that event A, or event B, or both occur is equal to the probability that event A occurs, plus the probability that event B occurs, minus the probability that the events occur simultaneously.*

*The addition rule may be written*

*P1A ́B2 = P1A2 + P1B2 - P1A ̈B2 (3.4.3)*

*When events A and B cannot occur simultaneously, P1A ̈ B2 is sometimes called “exclusive or,” and P1A ́ B2 = 0. When events A and B can occur simultaneously, P1A ́ B2 is sometimes called “inclusive or,” and we use the addition rule to calculate P1A ́ B2. Let us illustrate the use of the addition rule by means of an example.*

***EXAMPLE 3.4.6***

*If we select a person at random from the 318 subjects represented in Table 3.4.1, what is the probability that this person will be an Early age of onset subject 1E2 or will have no family history of mood disorders 1A2 or both?*

*Solution:*

*The probability we seek is P1E ́ A2. By the addition rule as expressed by Equation 3.4.3, this probability may be written as P1E ́ A 2 = P 1E 2 + P1A2 - P1E ̈ A2. We have already found that P1E2 = 141>318 = .4434 and P1E ̈ A2 = 28>318 = .0881. From the information in Table 3.4.1 we calculate P1A2 = 63>318 = .1981. Substituting these results into the equation for P1E ́A2 we have P1E ́A2 = .4434 + .1981 - .0881 =*

*Note that the 28 subjects who are both Early and have no family history of mood dis- orders are included in the 141 who are Early as well as in the 63 who have no family history of mood disorders. Since, in computing the probability, these 28 have been added into the numerator twice, they have to be subtracted out once to overcome the effect of duplication, or overlapping.*

***Independent Events*** *Suppose that, in Equation 3.4.2, we are told that event B has occurred, but that this fact has no effect on the probability of A. That is, suppose that the probability of event A is the same regardless of whether or not B occurs. In this situation, P1A ƒ B2 = P1A2. In such cases we say that A and B are independent events. The multiplication rule for two independent events, then, may be written as*

*P1A ̈ B2 = P1A2P1B2; P1A2 Z 0, P1B2 Z 0 (3.4.4)*

*Thus, we see that if two events are independent, the probability of their joint occur- rence is equal to the product of the probabilities of their individual occurrences.*

*Note that when two events with nonzero probabilities are independent, each of the following statements is true:*

*P1A ƒ B2 = P1A2, P1B ƒ A2 = P1B2, P1A ̈B2 = P1A2P1B2*

*Two events are not independent unless all these statements are true. It is important to be aware that the terms independent and mutually exclusive do not mean the same thing.*

*Let us illustrate the concept of independence by means of the following example.*

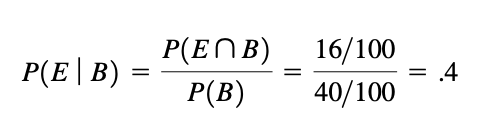
***EXAMPLE 3.4.7***

*In a certain high school class, consisting of 60 girls and 40 boys, it is observed that 24 girls and 16 boys wear eyeglasses. If a student is picked at random from this class, the probability that the student wears eyeglasses, P(E), is 40􏰉100, or .4.*

*(a) What is the probability that a student picked at random wears eyeglasses, given that the student is a boy?*

*Solution:*

*By using the formula for computing a conditional probability, we find this to be*

**

*Thus the additional information that a student is a boy does not alter the probability that the student wears eyeglasses, and P1E2 = P1E ƒ B2. We say that the events being a boy and wearing eyeglasses for this group are inde- pendent. We may also show that the event of wearing eyeglasses, E, and not being a boy, B are also independent as follows:*

*P1E ̈ B 2 24>100*

*P1E ƒ B2 = P1B2 = 60>100 = 60 = .4*

*(b) What is the probability of the joint occurrence of the events of wearing eyeglasses and being a boy?*

*Solution:*

*P1E ̈ B2 = P1B2P1E ƒ B2*

*but, since we have shown that events E and B are independent we may*

*3.4 CALCULATING THE PROBABILITY OF AN EVENT 75 Using the rule given in Equation 3.4.1, we have*

*replace P1E ƒ B2 by P1E2 to obtain, by Equation 3.4.4, P1E ̈ B2 = P1B2P1E2*

*= a 40 b a 40 b 100 100*

*= .16 s*

***Complementary Events*** *Earlier, using the data in Table 3.4.1, we computed the probability that a person picked at random from the 318 subjects will be an Early age of onset subject as P1E2 = 141>318 = .4434. We found the probability of a Later age at onset to be P1L2 = 177>318 = .5566. The sum of these two probabilities we found to be equal to 1. This is true because the events being Early age at onset and being Later age at onset are complementary events. In general, we may make the following statement about complemen- tary events. The probability of an event A is equal to 1 minus the probability of its comple- ment, which is written A, and*

**

*This follows from the third property of probability since the event, A, and its complement, A, are mutually exclusive.*

*EXAMPLE 3.4.8*

*Suppose that of 1200 admissions to a general hospital during a certain period of time, 750 are private admissions. If we designate these as set A, then A is equal to 1200 minus 750, or 450. We may computeP1A2 = 750>1200 = .625*

*P1A2 = 450>1200 = .375*

*P1A2 = 1 - P1A2*

*.375 = 1 - .625*

*.375 = .375 s*

***Marginal Probability***

*Earlier we used the term marginal probability to refer to a probability in which the numerator of the probability is a marginal total from a table such as Table 3.4.1. For example, when we compute the probability that a person picked at random from the 318 persons represented in Table 3.4.1 is an Early age of onset subject, the numerator of the probability is the total number of Early subjects, 141. Thus, P1E2 = 141>318 = .4434. We may define marginal probability more generally as follows:*

*DEFINITION*

*Given some variable that can be broken down into m categories designated by A , A , ..., A, ..., A and another jointly occurring 12im variable that is broken down into n categories designated b B , B , ..., B, ..., B , the marginal probability of A,P1A2, is*

*12jn ii*

*equal to the sum of the joint probabilities of Ai with all the cate-*

*gories of B. That is,*

*P1A2 􏰍 gP1A oB2, forallvaluesof j (3.4.6)*

*The following example illustrates the use of Equation 3.4.6 in the calculation of a marginal probability.*

***EXAMPLE 3.4.9***

*We wish to use Equation 3.4.6 and the data in Table 3.4.1 to compute the marginal prob- ability P(E).*

*Solution:*

*The variable age at onset is broken down into two categories, Early for onset 18 years or younger (E) and Later for onset occurring at an age over 18 years (L). The variable family history of mood disorders is broken down into four categories: negative family history (A), bipolar disorder only (B), unipolar disorder only (C ), and subjects with a history of both unipolar and bipolar disorder (D). The category Early occurs jointly with all four cate- gories of the variable family history of mood disorders. The four joint prob- abilities that may be computed are*

*P1E ̈ A2 = 28>318 = .0881 P1E ̈ B2 = 19>318 = .0597 P1E ̈ C2 = 41>318 = .1289 P1E ̈ D2 = 53>318 = .1667*

*We obtain the marginal probability P(E) by adding these four joint proba- bilities as follows:*

*P1E2 = P1E ̈A2 + P1E ̈B2 + P1E ̈C2 + P1E ̈D2 = .0881 + .0597 + .1289 + .1667*

*= .4434 s*

*The result, as expected, is the same as the one obtained by using the marginal total for Early as the numerator and the total number of subjects as the denominator.*

***EXERCISES***

*3.4.1 In a study of violent victimization of women and men, Porcerelli et al. (A-2) collected information from 679 women and 345 men aged 18 to 64 years at several family practice centers in the met- ropolitan Detroit area. Patients filled out a health history questionnaire that included a question about victimization. The following table shows the sample subjects cross-classified by sex and the type of violent victimization reported. The victimization categories are defined as no victimization, part- ner victimization (and not by others), victimization by persons other than partners (friends, family members, or strangers), and those who reported multiple victimization.*

*(a) Suppose we pick a subject at random from this group. What is the probability that this subject will be a woman?*

*(b) What do we call the probability calculated in part a?*

*(c) Show how to calculate the probability asked for in part a by two additional methods.*

*(d) If we pick a subject at random, what is the probability that the subject will be a woman and have experienced partner abuse?*

*(e) What do we call the probability calculated in part d?*

*(f) Suppose we picked a man at random. Knowing this information, what is the probability that he experienced abuse from nonpartners?*

*(g) What do we call the probability calculated in part f?*

*(h) Suppose we pick a subject at random. What is the probability that it is a man or someone who experienced abuse from a partner?*

*(i) What do we call the method by which you obtained the probability in part h?*

*3.4.2 Fernando et al. (A-3) studied drug-sharing among injection drug users in the South Bronx in New York City. Drug users in New York City use the term “split a bag” or “get down on a bag” to refer to the practice of dividing a bag of heroin or other injectable substances. A com- mon practice includes splitting drugs after they are dissolved in a common cooker, a procedure with considerable HIV risk. Although this practice is common, little is known about the preva- lence of such practices. The researchers asked injection drug users in four neighborhoods in the South Bronx if they ever “got down on” drugs in bags or shots. The results classified by gender and splitting practice are given below:*

*(a) How many marginal probabilities can be calculated from these data? State each in probabil- ity notation and do the calculations.*

*(b) How many joint probabilities can be calculated? State each in probability notation and do the calculations.*

*(c) How many conditional probabilities can be calculated? State each in probability notation and do the calculations.*

*(d) Use the multiplication rule to find the probability that a person picked at random never split drugs and is female.*

*(e) What do we call the probability calculated in part d?*

*(f) Use the multiplication rule to find the probability that a person picked at random is male, given that he admits to splitting drugs.*

*(g) What do we call the probability calculated in part f?*

*3.4.3 Refer to the data in Exercise 3.4.2. State the following probabilities in words and calculate: (a) P(Male ̈ Split Drugs)*

*(b) P(Male ́ Split Drugs)*

*(c) P(Male ƒ Split Drugs)*

*(d) P(Male)*

*2.4.4 Laveist and Nuru-Jeter (A-4) conducted a study to determine if doctor–patient race concordance was associated with greater satisfaction with care. Toward that end, they collected a national sample of African-American, Caucasian, Hispanic, and Asian-American respondents. The following table clas- sifies the race of the subjects as well as the race of their physician:*

*(a) What is the probability that a randomly selected subject will have an Asian/Pacific-Islander physician?*

* *(b) What is the probability that an African-American subject will have an African-American physician?*
* *(c) What is the probability that a randomly selected subject in the study will be Asian-American and have an Asian/Pacific-Islander physician?*

*(d) What is the probability that a subject chosen at random will be Hispanic or have a Hispanic physician?*

*(e) Use the concept of complementary events to find the probability that a subject chosen at random in the study does not have a white physician.   
3.4.5 If the probability of left-handedness in a certain group of people is .05, what is the probability of right-handedness (assuming no ambidexterity)?*

* *3.4.6 The probability is .6 that a patient selected at random from the current residents of a certain hos- pital will be a male. The probability that the patient will be a male who is in for surgery is .2. A patient randomly selected from current residents is found to be a male; what is the probability that the patient is in the hospital for surgery?*
* *3.4.7 In a certain population of hospital patients the probability is .35 that a randomly selected patient will have heart disease. The probability is .86 that a patient with heart disease is a smoker. What is the probability that a patient randomly selected from the population will be a smoker and have heart disease?*

***3.5 BAYES’ THEOREM, SCREENING TESTS, SENSITIVITY, SPECIFICITY, AND PREDICTIVE VALUE POSITIVE AND NEGATIVE***

*In the health sciences field a widely used application of probability laws and concepts is found in the evaluation of screening tests and diagnostic criteria. Of interest to clini- cians is an enhanced ability to correctly predict the presence or absence of a particular disease from knowledge of test results (positive or negative) and/or the status of present- ing symptoms (present or absent). Also of interest is information regarding the likeli- hood of positive and negative test results and the likelihood of the presence or absence of a particular symptom in patients with and without a particular disease.*

*In our consideration of screening tests, we must be aware of the fact that they are not always infallible. That is, a testing procedure may yield a false positive or a false negative.*

***DEFINITIONS***

1. *A false positive results when a test indicates a positive status when the true status is negative.*
2. *A false negative results when a test indicates a negative status when the true status is positive.*

*In summary, the following questions must be answered in order to evaluate the usefulness of test results and symptom status in determining whether or not a subject has some disease:*

1. *Given that a subject has the disease, what is the probability of a positive test result (or the presence of a symptom)?*
2. *Given that a subject does not have the disease, what is the probability of a negative test result (or the absence of a symptom)?*
3. *Given a positive screening test (or the presence of a symptom), what is the prob- ability that the subject has the disease?*
4. *Given a negative screening test result (or the absence of a symptom), what is the probability that the subject does not have the disease?*

*Table 3.5.1 Sample of n Subjects (Where n Is Large) Cross-Classified According to Disease Status and Screening Test Result*

*Suppose we have for a sample of n subjects (where n is a large number) the infor- mation shown in Table 3.5.1. The table shows for these n subjects their status with regard to a disease and results from a screening test designed to identify subjects with the dis- ease. The cell entries represent the number of subjects falling into the categories defined by the row and column headings. For example, a is the number of subjects who have the disease and whose screening test result was positive.*

*As we have learned, a variety of probability estimates may be computed from the information displayed in a two-way table such as Table 3.5.1. For example, we may compute the conditional probability estimate P1T ƒ D2 = a>1a + c2. This ratio is an estimate of the sensitivity of the screening test.*

***DEFINITION***

*The sensitivity of a test (or symptom) is the probability of a positive test result (or presence of the symptom) given the presence of the disease.*

*We may also compute the conditional probability estimate P1T ƒ D2 = d>1b + d2. This ratio is an estimate of the specificity of the screening test.*

***DEFINITION***

*The specificity of a test (or symptom) is the probability of a negative test result (or absence of the symptom) given the absence of the disease.*

*From the data in Table 3.5.1 we answer Question 3 by computing the conditional probability estimate P1D ƒ T 2. This ratio is an estimate of a probability called the pre- dictive value positive of a screening test (or symptom).*

***DEFINITION***

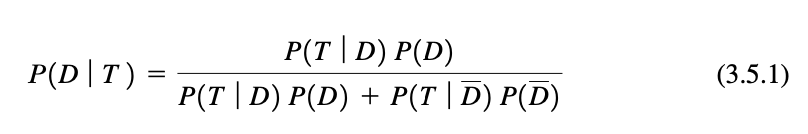
*The predictive value positive of a screening test (or symptom) is the probability that a subject has the disease given that the subject has a positive screening test result (or has the symptom).*

*Similarly, the ratio P1D ƒ T 2 is an estimate of the conditional probability that a subject does not have the disease given that the subject has a negative screening test result (or does not have the symptom). The probability estimated by this ratio is called the predictive value negative of the screening test or symptom.*

***DEFINITION***

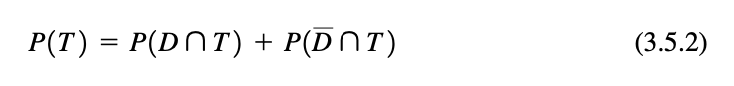
*The predictive value negative of a screening test (or symptom) is the probability that a subject does not have the disease, given that the subject has a negative screening test result (or does not have the symptom).*

*Estimates of the predictive value positive and predictive value negative of a test (or symptom) may be obtained from knowledge of a test’s (or symptom’s) sensitivity and specificity and the probability of the relevant disease in the general population. To obtain these predictive value estimates, we make use of Bayes’s theorem. The following statement of Bayes’s theorem, employing the notation established in Table 3.5.1, gives the predictive value positive of a screening test (or symptom):*

**

*It is instructive to examine the composition of Equation 3.5.1. We recall from Equation 3.4.2 that the conditional probability P1D ƒ T2 is equal to P1D ̈T2>P1T2. To understand the logic of Bayes’s theorem, we must recognize that the numerator of Equation 3.5.1 represents P1D ̈T2 and that the denominator represents P1T2. We know from the multiplication rule of probability given in Equation 3.4.1 that the numer- ator of Equation 3.5.1, P1T ƒ D2 P1D2, is equal to P1D ̈ T 2.*

*Now let us show that the denominator of Equation 3.5.1 is equal to P1T2. We know that event T is the result of a subject’s being classified as positive with respect to a screen- ing test (or classified as having the symptom). A subject classified as positive may have the disease or may not have the disease. Therefore, the occurrence of T is the result of a subject having the disease and being positive 3P1D ̈ T 24 or not having the disease and being positive 3P1D ̈ T 24. These two events are mutually exclusive (their intersec- tion is zero), and consequently, by the addition rule given by Equation 3.4.3, we may write*

* Since, by the multiplication rule, P1D ̈ T2 = P1T ƒ D2 P1D2 and P1D ̈ T2 =*

*P1T ƒ D2 P1D2, we may rewrite Equation 3.5.2 as*

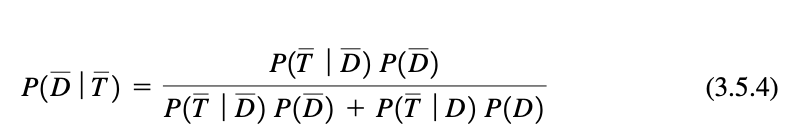
*P1T2 = P1T ƒ D2 P1D2 + P1T ƒ D2 P1D2 (3.5.3)*

*which is the denominator of Equation 3.5.1.*

*Note, also, that the numerator of Equation 3.5.1 is equal to the sensitivity times the rate (prevalence) of the disease and the denominator is equal to the sensitivity times the rate of the disease plus the term 1 minus the sensitivity times the term 1 minus the rate of the disease. Thus, we see that the predictive value positive can be calculated from knowl- edge of the sensitivity, specificity, and the rate of the disease.*

*Evaluation of Equation 3.5.1 answers Question 3. To answer Question 4 we follow a now familiar line of reasoning to arrive at the following statement of Bayes’s*

*theorem:*

**

*Equation 3.5.4 allows us to compute an estimate of the probability that a subject who is negative on the test (or has no symptom) does not have the disease, which is the predic- tive value negative of a screening test or symptom.*

*We illustrate the use of Bayes’ theorem for calculating a predictive value positive with the following example.*

***EXAMPLE 3.5.1***

*A medical research team wished to evaluate a proposed screening test for Alzheimer’s dis- ease. The test was given to a random sample of 450 patients with Alzheimer’s disease and an independent random sample of 500 patients without symptoms of the disease. The two samples were drawn from populations of subjects who were 65 years of age or older. The results are as follows:*

*Alzheimer’s Diagnosis?*

*P1DƒT2= P1T ƒD2P1D2 (3.5.4) P1T ƒ D2P1D2 + P1T ƒ D2P1D2*

*Test Result*

*Yes (D) 436*

*14*

*450*

*No (D)*

*5 441*

*495 509*

*500 950*

*Positive 1T2 Negative 1T2*

*Total*

*Total*

*Using these data we estimate the sensitivity of the test to be P1T ƒ D2 = 436>450 = .97. The specificity of the test is estimated to be P1T ƒ D2 = 495>500 = .99. We now use the results of the study to compute the predictive value positive of the test. That is, we wish to estimate the probability that a subject who is positive on the test has Alzheimer’s disease. From the tabulated data we compute P1T ƒ D2 = 436>450 = .9689 and P1T ƒ D2 = 5>500 = .01. Substitution of these results into Equation 3.5.1 gives*

*P1D ƒ T 2 = 1.96892 P1D2 (3.5.5)*

*We see that the predictive value positive of the test depends on the rate of the disease in the relevant population in general. In this case the relevant population consists of subjects who are 65 years of age or older. We emphasize that the rate of disease in the relevant general population, P1D2, cannot be computed from the sample data, since two independ- ent samples were drawn from two different populations. We must look elsewhere for an estimate of P1D2. Evans et al. (A-5) estimated that 11.3 percent of the U.S. population aged 65 and over have Alzheimer’s disease. When we substitute this estimate of P1D2 into Equation 3.5.5 we obtain 1.968921.1132*

*P1D ƒ T2 = 1.968921.1132 + 1.01211 - .1132 = .93*

*we see, in this case, the predictive value of the test is very high.*

*Similarly, let us now consider the predictive value negative of the test. We have already calculated all entries necessary except for P1T ƒ D2 = 14>450 = .0311. Using*

*s*

*the values previously obtained and our new value, we find*

*P1D ƒ T2 = (.99)(1 - .113) = .996 (.99)(1 - .113) + (.0311)(.113)*

*As we see, the predictive value negative is also quite high.*

*EXERCISES*

*3.5.1 A medical research team wishes to assess the usefulness of a certain symptom (call it S) in the diagnosis of a particular disease. In a random sample of 775 patients with the disease, 744 reported having the symptom. In an independent random sample of 1380 subjects without the disease, 21 reported that they had the symptom.*

* *(a) In the context of this exercise, what is a false positive?*
* *(b) What is a false negative?*
* *(c) Compute the sensitivity of the symptom.*
* *(d) Compute the specificity of the symptom.*
* *(e) Suppose it is known that the rate of the disease in the general population is .001. What is the*

*predictive value positive of the symptom?*

*(f) What is the predictive value negative of the symptom?*

*(g) Find the predictive value positive and the predictive value negative for the symptom for the following hypothetical disease rates: .0001, .01, and .10.*

*(h) What do you conclude about the predictive value of the symptom on the basis of the results obtained in part g?*

*3.5.2 In an article entitled “Bucket-Handle Meniscal Tears of the Knee: Sensitivity and Specificity of MRI signs,” Dorsay and Helms (A-6) performed a retrospective study of 71 knees scanned by MRI. One of the indicators they examined was the absence of the “bow-tie sign” in the MRI as evidence of a bucket-handle or “bucket-handle type” tear of the meniscus. In the study, surgery*

*confirmed that 43 of the 71 cases were bucket-handle tears. The cases may be cross-classified by “bow-tie sign” status and surgical results as follows:*

*Tear Surgically*

*Confirmed (D) 38*

*5*

*(a) What is the sensitivity of testing to see if the absent bow-tie sign indicates a meniscal tear? (b) What is the specificity of testing to see if the absent bow-tie sign indicates a meniscal tear?*

*(c) What additional information would you need to determine the predictive value of the test?*

*Oexle et al. (A-7) calculated the negative predictive value of a test for carriers of X-linked ornithine transcarbamylase deficiency (OTCD—a disorder of the urea cycle). A test known as the “allopuri- nol test” is often used as a screening device of potential carriers whose relatives are OTCD patients. They cited a study by Brusilow and Horwich (A-8) that estimated the sensitivity of the allopuri- nol test as .927. Oexle et al. themselves estimated the specificity of the allopurinol test as .997. Also they estimated the prevalence in the population of individuals with OTCD as 1􏰉32000. Use this information and Bayes’s theorem to calculate the predictive value negative of the allopurinol screening test.*

*Tear Surgically Confirmed As*

*Not Present (D)*

*10 48*

*18 23*

*Total*

*Positive Test*

*(absent bow-tie sign) 1T2*

*Negative Test*

*(bow-tie sign present) 1T2*

*43*

*and Specificity of MRI Signs,” Skeletal Radiology, 32 (2003), 266–272.*

*Total*

*Source: Theodore A. Dorsay and Clyde A. Helms, “Bucket-Handle Meniscal Tears of the Knee: Sensitivity*

*28 71*

*3.5.3*

*3.6 SUMMARY*

*In this chapter some of the basic ideas and concepts of probability were presented. The objective has been to provide enough of a “feel” for the subject so that the probabilis- tic aspects of statistical inference can be more readily understood and appreciated when this topic is presented later.*

*We defined probability as a number between 0 and 1 that measures the likelihood of the occurrence of some event. We distinguished between subjective probability and objective probability. Objective probability can be categorized further as classical or rel- ative frequency probability. After stating the three properties of probability, we defined and illustrated the calculation of the following kinds of probabilities: marginal, joint, and conditional. We also learned how to apply the addition and multiplication rules to find certain probabilities. We learned the meaning of independent, mutually exclusive, and complementary events. We learned the meaning of specificity, sensitivity, predic- tive value positive, and predictive value negative as applied to a screening test or disease symptom. Finally, we learned how to use Bayes’s theorem to calculate the probability that a subject has a disease, given that the subject has a positive screening test result (or has the symptom of interest).*

*REVIEW QUESTIONS AND EXERCISES*

1. *Define the following: (a) Probability  
      
   (c) Subjective probability  
    (e) The relative frequency concept of probability (g) Independence  
      
   (i) Joint probability (k) The addition rule  
      
   (m) Complementary events (o) False negative  
    (q) Specificity  
      
   (s) Predictive value negative*
2. *Name and explain the three properties of probability.*

*(b) Objective probability (d) Classical probability*

*(f) Mutually exclusive events (h) Marginal probability*

*(j) Conditional probability*

*(l) The multiplication rule (n) False positive*

*(p) Sensitivity*

*(r) Predictive value positive (t) Bayes’s theorem*

*3. Coughlin et al. (A-9) examined the breast and cervical screening practices of Hispanic and non-Hispanic women in counties that approximate the U.S. southern border region. The study used data from the Behavioral Risk Factor Surveillance System surveys of adults age 18 years or older conducted in 1999 and 2000. The table below reports the number of observations of Hispanic and non-Hispanic women who had received a mammogram in the past 2 years cross-classified with marital status.*

*(a) We select at random a subject who had a mammogram. What is the probability that she is*

*divorced or separated?*

*(b) We select at random a subject who had a mammogram and learn that she is Hispanic. With that information, what is the probability that she is married?*

*(c) We select at random a subject who had a mammogram. What is the probability that she is non-Hispanic and divorced or separated?*

*(d) We select at random a subject who had a mammogram. What is the probability that she is Hispanic or she is widowed?*

*(e) We select at random a subject who had a mammogram. What is the probability that she is not married?*

*4. Swor et al. (A-10) looked at the effectiveness of cardiopulmonary resuscitation (CPR) training in people over 55 years old. They compared the skill retention rates of subjects in this age group who completed a course in traditional CPR instruction with those who received chest-compression only cardiopulmonary resuscitation (CC-CPR). Independent groups were tested 3 months after training. The table below shows the skill retention numbers in regard to overall competence as assessed by video ratings done by two video evaluators.*

*Rated Overall*

*Competent CPR CC-CPR Total*

*Yes 12 15 27 No 15 14 29*

*Total 27 29 56*

*Source: Robert Swor, Scott Compton, Fern Vining, Lynn Ososky Farr, Sue Kokko, Rebecca Pascual, and Raymond E. Jackson, “A Randomized Controlled Trial of Chest Compression Only CPR for Older Adults—a Pilot Study,” Resuscitation, 58 (2003), 177–185.*

* *(a) Find the following probabilities and explain their meaning:* 
  1. *A randomly selected subject was enrolled in the CC-CPR class.*
  2. *A randomly selected subject was rated competent.*
  3. *A randomly selected subject was rated competent and was enrolled in the CPR course.*
  4. *A randomly selected subject was rated competent or was enrolled in CC-CPR.*
  5. *A randomly selected subject was rated competent given that he or she enrolled in the  
        
     CC-CPR course.*
* *(b) We define the following events to be*

*A = a subject enrolled in the CPR course*

*B = a subject enrolled in the CC-CPR course*

*C = a subject was evaluated as competent*

*D = a subject was evaluated as not competent*

*Then explain why each of the following equations is or is not a true statement:*

*1. P1A ̈C2 = P1C ̈A2*

*3. P1A2 = P1A ́C2 + P1A ́D2 5. P1D ƒ A2 = P1D2*

*7. P1A ̈B2 = 0*

*9. P1A ̈D2 = P1A2P1A ƒ D2*

*2. P1A ́B2 = P1B ́A2*

*4. P1B ́C2 = P1B2 + P1C2 6. P1C ̈B2 = P1C2P1B2*

*8. P1C ̈B2 = P1B2P1C ƒ B2*

*5.*

*SOME BASIC PROBABILITY CONCEPTS*

*Pillmann et al. (A-11) studied patients with acute brief episodes of psychoses. The researchers classified subjects into four personality types: obsessoid, asthenic/low self-confident, asthenic/high self-confident, nervous/tense, and undeterminable. The table below cross-classifies these personal- ity types with three groups of subjects—those with acute and transient psychotic disorders (ATPD), those with “positive” schizophrenia (PS), and those with bipolar schizo-affective disorder (BSAD):*

*6.*

*7.*

*Personality Type*

*Obsessoid (O)*

*Asthenic/low Self-confident (A) Asthenic/high Self-confident (S) Nervous/tense (N ) Undeterminable (U )*

*ATPD (1)*

*9 20 5 4 4*

*PS (2)*

*2 17 3 7 13*

*42*

*BSAD (3)*

*Total*

*Total 42*

*6 17 15 52 8 16 4 15 9 26*

*42 126*

*Source: Frank Pillmann, Raffaela Blõink, Sabine Balzuweit, Annette Haring, and Andreas Marneros, “Personality and Social Interactions in Patients with Acute Brief Psychoses,” Journal of Nervous and Mental Disease, 191 (2003), 503–508.*

*(a) P1O2 (b) P1A ́ 22 (c) P112 (d) P1A2 (e) P1A ƒ 32 (f) P132 (g) P12 ̈32 (h) P12 ƒ A2*

*A certain county health department has received 25 applications for an opening that exists for a public health nurse. Of these applicants 10 are over 30 and 15 are under 30. Seventeen hold bach- elor’s degrees only, and eight have master’s degrees. Of those under 30, six have master’s degrees. If a selection from among these 25 applicants is made at random, what is the probability that a person over 30 or a person with a master’s degree will be selected?*

*The following table shows 1000 nursing school applicants classified according to scores made on a college entrance examination and the quality of the high school from which they graduated, as rated by a group of educators:*

*Quality of High Schools*

*Poor Average Superior*

*Score (P) (A) (S) Total*

*Find the following probabilities if a subject in this study is chosen at random:*

*Low (L) Medium (M) High (H)*

*105 60 55 220 70 175 145 390 25 65 300 390*

*Total 200 300 500 1000*

*(a) Calculate the probability that an applicant picked at random from this group:*

1. *Made a low score on the examination.*
2. *Graduated from a superior high school.*
3. *Made a low score on the examination and graduated from a superior high school.*
4. *Made a low score on the examination given that he or she graduated from a superior high  
      
   school.*
5. *Made a high score or graduated from a superior high school.*

*b) Calculate the following probabilities:*

*4. P1A ƒ H2 5. P1M ̈P2 6. 1H ƒ S2*

1. *If the probability that a public health nurse will find a client at home is .7, what is the probabil- ity (assuming independence) that on two home visits made in a day both clients will be home?*
2. *For a variety of reasons, self-reported disease outcomes are frequently used without verification in epi- demiologic research. In a study by Parikh-Patel et al. (A-12), researchers looked at the relationship between self-reported cancer cases and actual cases. They used the self-reported cancer data from a California Teachers Study and validated the cancer cases by using the California Cancer Registry data. The following table reports their findings for breast cancer:*

*Cancer Reported (A) Yes*

*No*

*Total*

*Cancer in Registry (B) 2991*

*112*

*3103*

*Cancer Not in Registry*

*2244 115849*

*118093*

*Total*

*5235 115961*

*121196*

*Source: Arti Parikh-Patel, Mark Allen, William E. Wright, and the California Teachers Study Steering Committee, “Validation of Self-Reported Cancers in the California Teachers Study,” American Journal of Epidemiology, 157 (2003), 539–545.*

*(a) Let A be the event of reporting breast cancer in the California Teachers Study. Find the prob- ability of A in this study.*

*(b) Let B be the event of having breast cancer confirmed in the California Cancer Registry. Find the probability of B in this study.*

*(c) Find P1A ̈ B2*

*(d) Find P1A ƒ B2*

*(e) Find P1B ƒ A2*

*(f) Find the sensitivity of using self-reported breast cancer as a predictor of actual breast cancer in the California registry.*

*(g) Find the specificity of using self-reported breast cancer as a predictor of actual breast cancer in the California registry.*

1. *In a certain population the probability that a randomly selected subject will have been exposed to a certain allergen and experience a reaction to the allergen is .60. The probability is .8 that a subject exposed to the allergen will experience an allergic reaction. If a subject is selected at ran- dom from this population, what is the probability that he or she will have been exposed to the allergen?*
2. *Suppose that 3 percent of the people in a population of adults have attempted suicide. It is also known that 20 percent of the population are living below the poverty level. If these two events are independent, what is the probability that a person selected at random from the population will have attempted suicide and be living below the poverty level?*
3. *In a certain population of women 4 percent have had breast cancer, 20 percent are smokers, and 3 percent are smokers and have had breast cancer. A woman is selected at random from the popu- lation. What is the probability that she has had breast cancer or smokes or both?*
4. *The probability that a person selected at random from a population will exhibit the classic symp- tom of a certain disease is .2, and the probability that a person selected at random has the disease*

*is .23. The probability that a person who has the symptom also has the disease is .18. A person selected at random from the population does not have the symptom. What is the probability that the person has the disease?*

*Refer to Exercise 14. State in words the event E = 1A ́ B2. Refer to Exercise 14. State in words the event F = 1B ́ C 2. Refer to Exercise 14. Comment on the event G = 1A ̈ B2.*

*For a certain population we define the following events with respect to plasma lipoprotein levels (mg􏰊dl): A = (10–15); B = 1Ú302;C = 1...202. Are the events A and B mutually exclusive? A and C? B and C? Explain your answer to each question.*

*Refer to Exercise 18. State in words the meaning of the following events:*

*(a) A ́B (b) A ̈B (c) A ̈C (d) A ́C*

*Refer to Exercise 18. State in words the meaning of the following events:*

*(a) A (b) B (c) C*

*Rothenberg et al. (A-13) investigated the effectiveness of using the Hologic Sahara Sonometer, a portable device that measures bone mineral density (BMD) in the ankle, in predicting a fracture. They used a Hologic estimated bone mineral density value of .57 as a cutoff. The results of the investigation yielded the following data:*

*Present 1D2 214*

*73*

*287*

*Not Present 1D2 Total 670 884*

*330 403*

*1000 1287*

*Confirmed Fracture*

*BMD = .571T2 BMD 7 .571T2*

*Total*

*22.*

*Source: Ralph J. Rothenberg, M.D., Joan L. Boyd, Ph.D., and John P. Holcomb, Ph.D. Used with permission.*

*(a) Calculate the sensitivity of using a BMD value of .57 as a cutoff value for predicting fracture and interpret your results.*

*(b) Calculate the specificity of using a BMD value of .57 as a cutoff value for predicting fracture and interpret your results.*

*Verma et al. (A-14) examined the use of heparin-PF4 ELISA screening for heparin-induced throm- bocytopenia (HIT) in critically ill patients. Using C-serotonin release assay (SRA) as the way of validating HIT, the authors found that in 31 patients tested negative by SRA, 22 also tested neg- ative by heparin-PF4 ELISA.*

*(a) Calculate the specificity of the heparin-PF4 ELISA testing for HIT.*

*(b) Using a “literature-derived sensitivity” of 95 percent and a prior probability of HIT occur-*

*rence as 3.1 percent, find the positive predictive value.*

*(c) Using the same information as part (b), find the negative predictive value.*

*23. The sensitivity of a screening test is .95, and its specificity is .85. The rate of the disease for which*

*the test is used is .002. What is the predictive value positive of the test?*

*Exercises for Use with Large Data Sets Available on the Following Website: www.wiley.com/college/daniel*

*Refer to the random sample of 800 subjects from the North Carolina birth registry we investigated in the Chapter 2 review exercises.*

1. *Create a table that cross-tabulates the counts of mothers in the classifications of whether the baby was premature or not (PREMIE) and whether the mother admitted to smoking during pregnancy (SMOKE) or not.  
    (a) Find the probability that a mother in this sample admitted to smoking.  
      
   (b) Find the probability that a mother in this sample had a premature baby.  
    (c) Find the probability that a mother in the sample had a premature baby given that the mother admitted to smoking.  
    (d) Find the probability that a mother in the sample had a premature baby given that the mother did not admit to smoking.  
    (e) Find the probability that a mother in the sample had a premature baby or that the mother did not admit to smoking.*
2. *Create a table that cross-tabulates the counts of each mother’s marital status (MARITAL) and whether she had a low birth weight baby (LOW).  
    (a) Find the probability a mother selected at random in this sample had a low birth weight baby. (b) Find the probability a mother selected at random in this sample was married.  
      
   (c) Find the probability a mother selected at random in this sample had a low birth weight child given that she was married.  
    (d) Find the probability a mother selected at random in this sample had a low birth weight child given that she was not married.  
      
   (e) Find the probability a mother selected at random in this sample had a low birth weight child and the mother was married.*

*REFERENCES*

*Methodology References*

1. *ALLAN GUT, An Intermediate Course in Probability, Springer-Verlag, New York, 1995.*
2. *RICHARD ISAAC, The Pleasures of Probability, Springer-Verlag, New York, 1995.*
3. *HAROLD J. LARSON, Introduction to Probability, Addison-Wesley, Reading, MA, 1995.*
4. *L. J. SAVAGE, Foundations of Statistics, Second Revised Edition, Dover, New York, 1972.*
5. *A. N. KOLMOGOROV, Foundations of the Theory of Probability, Chelsea, New York, 1964 (original German  
      
   edition published in 1933).  
      
   Applications References*

* *A-1. TASHA D. CARTER, EMANUELA MUNDO, SAGAR V. PARKh, and JAMES L. KENNEDY, “Early Age at Onset as a Risk Factor for Poor Outcome of Bipolar Disorder,” Journal of Psychiatric Research, 37 (2003), 297–303.*
* *A-2. JOHN H. PORCERELLI, ROSEMARY COGAN, PATRICIA P. WEST, EDWARD A. ROSE, DAWN LAMBRECHT, KAREN E. WILSON, RICHARD K. SEVERSON, and DUNIA KARANA,“Violent Victimization of Women and Men: Physi- cal and Psychiatric Symptoms,” Journal of the American Board of Family Practice, 16 (2003), 32–39.*

*A-3.*

*A-4. A-5.*

*A-6. A-7.*

*A-8.*

*A-9. A-10. A-11. A-12. A-13. A-14.*

*SOME BASIC PROBABILITY CONCEPTS*

*DANIEL FERNANDO, ROBERT F. SCHILLING, JORGE FONTDEVILA, and NABILA EL-BASSEL, “Predictors of Shar- ing Drugs Among Injection Drug Users in the South Bronx: Implications for HIV Transmission,” Journal of Psychoactive Drugs, 35 (2003), 227–236.*

*THOMAS A. LAVEIST and AMANI NURU-JETER, “Is Doctor–Patient Race Concordance Associated with Greater Satisfaction with Care?” Journal of Health and Social Behavior, 43 (2002), 296–306.*

*D. A. EVANS, P. A. SCHERR, N. R. COOK, M. S. ALBERT, H. H. FUNKENSTEIN, L. A. SMITH, L. E. HEBERT, T. T. WETLE, L. G. BRANCH, M. CHOWN, C. H. HENNEKENS, and J. O. TAYLOR, “Estimated Prevalence of Alzheimer’s Disease in the United States,” Milbank Quarterly, 68 (1990), 267–289.*

*THEODORE A. DORSAY and CLYDE A. HELMS,“Bucket-Handle Meniscal Tears of the Knee: Sensitivity and Specificity of MRI Signs,” Skeletal Radiology, 32 (2003), 266–272.*

*KONRAD OEXLE, LUISA BONAFE, and BEAT STENMANN, “Remark on Utility and Error Rates of the Allopuri- nol Test in Detecting Mild Ornithine Transcarbamylase Deficiency,” Molecular Genetics and Metabolism, 76 (2002), 71–75.*

*S. W. BRUSILOW, A. L. HORWICH, “Urea Cycle Enzymes,” in: C. R. SCRIVER, A. L. BEAUDET, W. S. SLY, D. VALLE (Eds.), The Metabolic and Molecular Bases of Inherited Disease, 8th ed., McGraw-Hill, New York, 2001, pp. 1909–1963.*

*STEVEN S. COUGHLIN, ROBERT J. UHLER, THOMAS RICHARDS, and KATHERINE M. WILSON, “Breast and Cer- vical Cancer Screening Practices Among Hispanic and Non-Hispanic Women Residing Near the United States–Mexico Border, 1999–2000,” Family and Community Health, 26 (2003), 130–139.*

*ROBERT SWOR, SCOTT COMPTON, FERN VINING, LYNN OSOSKY FARR, SUE KOKKO, REBECCA PASCUAL, and RAYMOND E. JACKSON, “A Randomized Controlled Trial of Chest Compression Only CPR for Older Adults— a Pilot Study,” Resuscitation, 58 (2003), 177–185.*

*FRANK PILLMANN, RAFFAELA BLÕINK, SABINE BALZUWEIT, ANNETTE HARING, and ANDREAS MARNEROS, “Personality and Social Interactions in Patients with Acute Brief Psychoses,” Journal of Nervous and Mental Disease, 191 (2003), 503–508.*

*ARTI PARIKH-PATEL, MARK ALLEN, WILLIAM E. WRIGHT, and the California Teachers Study Steering Com- mittee, “Validation of Self-Reported Cancers in the California Teachers Study,” American Journal of Epidemi- ology, 157 (2003), 539–545.*

*RALPH J. ROTHENBERG, JOAN L. BOYD, and JOHN P. HOLCOMB, “Quantitative Ultrasound of the Calcaneus as a Screening Tool to Detect Osteoporosis: Different Reference Ranges for Caucasian Women, African-American Women, and Caucasian Men,” Journal of Clinical Densitometry, 7 (2004), 101–110.*

*ARUN K. VERMA, MARC LEVINE, STEPHEN J. SHALANSKY, CEDRIC J. CARTER, and JOHN G. KELTON, “Fre- quency of Heparin-Induced Thrombocytopenia in Critical Care Patients,” Pharmacotheray, 23 (2003), 645–753.*