# 1 The Role of Statistics

In this chapter we informally discuss how statistics is used to attempt to answer questions raised in research. Because probability is basic to statistical decision making, we will also present a few probability rules to show how probabilities are computed. Since this is an overview, we make no attempt to give precise definitions. The more formal development will follow in later chapters.

## 1.1. THE BASIC STATISTICAL PROCEDURE

Scientists sometimes use statistics to describe the results of an experiment or an investigation. This process is referred to as *data analysis* or *descriptive statistics*. Scientists also use statistics another way; if the entire population of interest is not accessible to them for some reason, they often observe only a portion of the population (a sample) and use statistics to answer questions about the whole population. This process is called *inferential statistics*. Statistical inference is the main focus of this book.

Inferential statistics can be defined as the science of using *probability* to make decisions. Before explaining how this is done, a quick review of the "laws of chance" is in order. Only four probability rules will be discussed here, those for (1) simple probability, (2) mutually exclusive events, (3) independent events, and (4) conditional probability. For anyone wanting more than covered here, Johnson and Kuby (2000) as well as Bennett, Briggs, and Triola (2003) provide more detailed discussion.

Early study of probability was greatly influenced by games of chance. Wealthy games players consulted mathematicians to learn if their losses during a night of gaming were due to bad luck or because they did not know how to compute their chances of winning. (Of course, there was always the possibility of chicanery, but that seemed a matter better settled with dueling weapons than mathematical computations.) Stephen Stigler (1986) states that formal study of probability began in 1654 with the exchange of letters between two famous French mathematicians, Blaise Pascal and Pierre de Fermat, regarding a question posed by a French nobleman about a dice game. The problem can be found in Exercise 1.1.5.

In games of chance, as in experiments, we are interested in the outcomes of a random phenomenon that cannot be predicted with certainty because usually there is more than one outcome and each is subject to chance. The probability of an outcome is a measure of how likely that outcome is to occur. The random outcomes associated with games of chance should be equally likely to occur if the gambling device is *fair*, controlled by chance alone. Thus the probability of getting a head on a single toss of a fair coin and the probability of getting an even number when we roll a fair die are both $1/2$.

*Statistics for Research*, *Third Edition*, Edited by Shirley Dowdy, Stanley Weardon, and Daniel Chilko.
ISBN 0-471-26735-X 

Because of the early association between probability and games of chance, we label some collection of equally likely outcomes as a success. A collection of outcomes is called an event. If success is the event of an even number of pips on a fair die, then the event consists of outcomes 2, 4, and 6. An event may consist of only one outcome, as the event head on a single toss of a coin. The probability of a success is found by the following probability rule:

$$\text{probability of success} = \frac{\text{number of successful outcomes}}{\text{total number of outcomes}}$$

In symbols

$$P(\text{success}) = P(S) = \frac{n_s}{N}$$

where $n_S$ is the number of outcomes in the event designated as success and $N$ is the total number of possible outcomes. Thus the simple probability rule for equally likely outcomes is to count the number of ways a success can be obtained and divide it by the total number of outcomes.

---

**Example 1.1. Simple Probability Rule for Equally Likely Outcomes**

There is a game, often played at charity events, that involves tossing a coin such as a 25-cent piece. The quarter is tossed so that it bounces off a board and into a chute to land in one of nine glass tumblers, only one of which is red. If the coin lands in the red tumbler, the player wins \$1; otherwise the coin is lost. In the language of probability, there are $N = 9$ possible outcomes for the toss and only one of these can lead to a success. Assuming skill is not a factor in this game, all nine outcomes are equally likely and $P(\text{success}) = 1/9$.

In the game described above, $P(\text{win}) = 1/9$ and $P(\text{loss}) = 8/9$. We observe there is only one way to win \$1 and eight ways to lose 25¢. A related idea from the early history of probability is the concept of *odds*. The odds for winning are $P(\text{win})/P(\text{loss})$. Here we say, "The odds for winning are one to eight" or, more pessimistically, "The odds against winning are eight to one." In general,

$$\text{odds for success} = \frac{P(\text{success})}{1 - P(\text{success})}$$

---

We need to stress that the simple probability rule above applies only to an experiment with a discrete number of equally likely outcomes. There is a similarity in computing probabilities for *continuous variables* for which there is a distribution curve for measures of the variable. In this case

$$P(\text{success}) = \frac{\text{area under the curve where the measure is called a success}}{\text{total area under the curve}}$$

A simple example is provided by the "spinner" that comes with many board games. The spinner is an arrow that spins freely around an axle attached to the center of a circle. Suppose that the circle is divided into quadrants marked 1, 2, 3, and 4 and play on the board is determined by the quadrant in which the spinner comes to rest. If no skill is involved in spinning the arrow, the outcomes can be considered uniformly distributed over the 360° of the circle. If it is a success to land in the third quadrant of the circle, a spin is a success when the arrow stops anywhere in the 90° of the third quadrant and

$$P(\text{success}) = \frac{\text{area in third quadrant}}{\text{total area}} = \frac{90}{360} = \frac{1}{4}$$

While only a little geometry is needed to calculate probabilities for a *uniform distribution*, knowledge of calculus is required for more complex distributions. However, finding probabilities for many continuous variables is possible by using simple tables. This will be explained in later chapters.

The next rule involves events that are *mutually exclusive*, meaning one event excludes the possibility of another. For instance, if two dice are rolled and the event is that the sum of spots is $y = 7$, then $y$ cannot possibly be another value as well. However, there are six ways that the spots, or pips, on two dice can produce a sum of 7, and each of these is mutually exclusive of the others. To see how this is so, imagine that the pair consists of one red die and one green; then we can detail all the possible outcomes for the event $y = 7$:

| | | | | | | |
|---|---|---|---|---|---|---|
| Red die: | 1 | 2 | 3 | 4 | 5 | 6 |
| Green die: | 6 | 5 | 4 | 3 | 2 | 1 |
| Sum: | 7 | 7 | 7 | 7 | 7 | 7 |

If a success depends only on a value of $y = 7$, then by the simple probability rule the number of possible successes is $n_S = 6$; the number of possible outcomes is $N = 36$ because each of the six outcomes of the red die can be paired with each of the six outcomes of the green die and the total number of outcomes is $6 \times 6 = 36$. Thus $P(\text{success}) = n_S/N = 6/36 = 1/6$. However, we need a more general statement to cover mutually exclusive events, whether or not they are equally likely, and that is the addition rule.

If a success is any of $k$ mutually exclusive events $E_1, E_2, \ldots, E_k$, then the *addition rule for mutually exclusive events* is $P(\text{success}) = P(E_1) + P(E_2) + \cdots + P(E_k)$. This holds true with the dice; if $E_1$ is the event that the red die shows 1 and the green die shows 6, then $P(E_1) = 1/36$. Then, because each of the $k = 6$ events has the same probability,

$$P(\text{success}) = \left(\frac{1}{36}\right) + \left(\frac{1}{36}\right) + \left(\frac{1}{36}\right) + \left(\frac{1}{36}\right) + \left(\frac{1}{36}\right) + \left(\frac{1}{36}\right) = \frac{6}{36} = \frac{1}{6}$$

Here $1/36$ is the common probability for all events, but the addition rule for mutually exclusive events still holds true even when the probability values are not the same for all events.

---

**Example 1.2. Addition Rule for Mutually Exclusive Events**

To see how this rule applies to events that are not equally likely, suppose a coin-operated gambling device is programmed to provide, on random plays, winnings with the following probabilities:

| Event | $P$(Event) |
|---|---|
| Win 10 coins | 0.001 |
| Win 5 coins | 0.010 |

| Event | $P$(Event) |
|---|---|
| Win 3 coins | 0.040 |
| Win 1 coin | 0.359 |
| Lose 1 coin | 0.590 |

Because most players consider it a success if any coins are won, $P(\text{success}) = 0.0001 + 0.010 + 0.040 + 0.359 = 0.410$, and the odds for winning are $0.41/0.59 = 0.695$, while the odds against a win are $0.59/0.41 = 1.44$.

We might ask why we bother to add $0.0001 + 0.010 + 0.040 + 0.359$ to obtain $P(\text{success}) = 0.41$ when we can obtain it just from knowledge of $P$(no success). On a play at the coin machine, one either wins of loses, so there is the probability of a success, $P(S) = 0.41$, and the probability of no success, $P(\text{no success}) = 0.59$. The opposite of a success, is called its *complement*, and its probability is symbolized as $P(\bar{S})$. In a play at the machine there is no possibility of neither a win nor a loss, $P(S) + P(\bar{S}) = 1.0$, so rather than counting the four ways to win it is easier to find $P(S) = 1.0 - P(\bar{S}) = 1.0 - 0.59 = 0.41$. Note that in the computation of the odds for winning we used the ratio of the probability of a win to its complement, $P(S)/P(\bar{S})$.

---

At games of chance, people who have had a string of losses are encouraged to continue to play with such remarks as “Your luck is sure to change” or “Odds favor your winning now,” but is that so? Not if the plays, or events, are independent. A play in a game of chance has no memory of what happened on previous plays. So using the results of Example 1.2, suppose we try the machine three times. The probability of a win on the first play is $P(S_1) = 0.41$, but the second coin played has no memory of the fate of its predecessor, so $P(S_2) = 0.41$, and likewise $P(S_3) = 0.41$. Thus we could insert 100 coins in the machine and lose on the first 99 plays, but the probability that our last coin will win remains $P(S_{100}) = 0.41$. However, we would have good reason to suspect the honesty of the machine rather than bad luck, for with an honest machine for which the probability of a win is 0.41, we would expect about 41 wins in 100 plays.

When dealing with independent events, we often need to find the *joint probability* that two or more of them will all occur simultaneously. If the total number of possible outcomes ($N$) is small, we can always compile tables, so with the $N = 52$ cards in a standard deck, we can classify each card by color (red or black) and as to whether or not it is an honor card (ace, king, queen, or jack). Then we can sort and count the cards in each of four groups to get the following table:

| | Color | | |
|---|---|---|---|
| Honor | Black | Red | Total |
| No | 18 | 18 | 36 |
| Yes | 8 | 8 | 16 |
| Total | 26 | 26 | 52 |

If a card is dealt at random from such a deck, we can find the *joint probability* that it will be red and an honor by noting that there are 8 such cards in the deck of 52; hence $P(\text{red and honor}) = P(RH) = 8/52 = 2/13$. This is easy enough when the total number of outcomes is small or when they have already been tabulated, but in many cases there are too many or there is a process such as the slot machine capable of producing an infinite number of outcomes. Fortunately there is a probability rule for such situations.

*The multiplication rule* for finding the joint probability of $k$ independent events $E_1$, $E_2, \ldots, E_k$ is

$$P(E_1 \text{ and } E_2 \text{ and } \ldots E_k) = P(E_1) \times P(E_2) \times \cdots \times P(E_k)$$

With the cards, $k$ is 2, $E_1$ is a red card, and $E_2$ is an honor card, so $P(E_1E_2) = P(E_1) \times P(E_2) = (26/52) \times (16/52) = (1/2) \times (4/13) = 4/26 = 2/13$.

---

**Example 1.3. The Multiplication Rule for Independent Events**

Gender and handedness are independent, and if $P(\text{female}) = 0.50$ and $P(\text{left handed}) = 0.15$, then the probability that the first child of a couple will be a left-handed girl is

$$P(\text{female and left handed}) = P(\text{female}) \times P(\text{left handed}) = 0.50 \times 0.15 = 0.075$$

If the probability values $P$(female) and $P$(left handed) are realistic, the computation is easier than the alternative of trying to tabulate the outcomes of all first births. We know the biological mechanism for determining gender but not handedness, so it was only estimated here. However, the value we would obtain from a tabulation of a large number of births would also be only an estimate. We will see in Chapter 3 how to make estimates and how to say scientifically, "The probability that the first child will be a left-handed girl is likely somewhere around 0.075."

---

The multiplication rule is very convenient when events are independent, but frequently we encounter events that are not independent but rather are at least partially related. Thus we need to understand these and how to deal with them in probability. When told that a person is from Sweden or some other Nordic country, we might immediately assume that he or she has blue eyes, or conversely dark eyes if from a Mediterranean country. In our encounters with people from these areas, we think we have found that the probability of eye color $P$(blue) is not the same for both those geographic regions but rather depends, or *is conditioned*, on the region from which a person comes. *Conditional probability* is symbolized as $P(E_2|E_1)$, and we say "The probability of event 2 given event 1." In the case of eye color, it would be the probability of blue eyes given that one is from a Nordic country.

The *conditional probability rule* for finding the conditional probability of event 2 given event 1 is

$$P(E_2|E_1) = \frac{P(E_1E_2)}{P(E_1)}$$

In the deck of cards, the probability a randomly dealt card will be red and an honor card is $P(\text{red and honor}) = 8/52$, while the probability it is red is $P(R) = 26/52$, so the probability that it will be an honor card, given that it is a red card is $P(RH)/P(R) = 8/26 = 4/13$, which is the same as $P(H)$ because the two are independent rather than related. Hence independent events can be defined as satisfying $P(E_2|E_1) = P(E_2)$.

**Example 1.4. The Conditional Probability Rule**

Suppose an oncologist is suspicious that cancer of the gum may be associated with use of smokeless tobacco. It would be ideal if he also had data on the use of smokeless tobacco by those free of cancer, but the only data immediately available are from 100 of his own cancer patients, so he tabulates them to obtain the following:

Smokeless Tobacco

| Cancer Site | No | Yes | Total |
|---|---|---|---|
| Gum | 5 | 20 | 25 |
| Elsewhere | 60 | 15 | 75 |
| Total | 65 | 35 | 100 |

There are 25 cases of gum cancer in his database and 20 of those patients had used smokeless tobacco, so we see that his best estimate of the probability that a randomly drawn gum cancer patient was a user of smokeless tobacco is $20/25 = 0.80$. This probability could also be found by the conditional probability rule. If $P(\text{gum}) = P(G)$ and $P(\text{user}) = P(U)$, then

$$P(U|G) = \frac{P(GU)}{P(G)} = \frac{(20/100)}{(25/100)} = \frac{20}{25} = 0.80$$

Are gum cancer and use of smokeless tobacco independent? They are if $P(U|G) = P(U)$, and from the data set, the best estimate of users among all cancer patients is $P(U) = 35/100 = 0.35$. The discrepancy in estimates is 0.80 for gum cancer patients compared to 0.35 for all patients. This leads us to believe that gum cancer and smokeless tobacco usage are related rather than independent. In Chapter 5, we will see how to test to see whether or not two variables are independent.

Odds obtained from medical data sets similar to but much larger than that in Example 1.4 are frequently cited in the news. Had the odds been the same in a data set of hundreds or thousands of gum cancer patients, we would report that the odds were $0.80/0.20 = 4.0$ for smokeless tobacco, and $0.35/0.65 = 0.538$ for smokeless tobacco among all cancer patients. Then, for sake of comparison, we would report the *odds ratio*, which is the ratio of the two odds, $4.0/0.538 = 7.435$. This ratio gives the relative frequency of smokeless tobacco users among gum cancer patients to smokeless tobacco users among all cancer patients, and the medical implications are ominous. For comparison, it would be helpful to have data on the usage of smokeless tobacco in a cancer-free population, but first information about an association such as that in Example 1.4 usually comes from medical records for those with a disease.

Caution is necessary when trying to interpret odds ratios, especially those based on very low incidences of occurrence. To show a totally meaningless odds ratio, suppose we have two data sets, one containing 20 million broccoli eaters and the other of 10 million who do not eat the vegetable. Then, if we examine the health records of those in each group, we find there are two in each group suffering from chronic bladder infections. The odds ratio is 2.0, but we would garner strange looks rather than prestige if we attempted to claim that the odds for

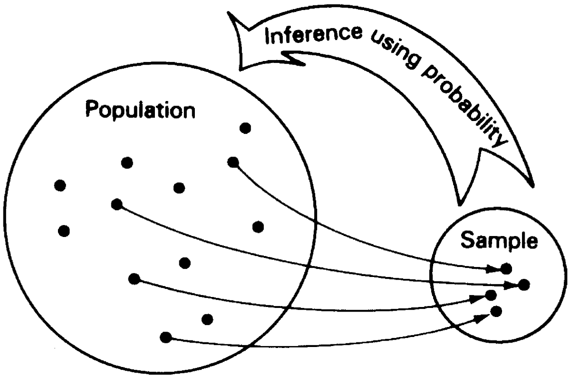


**FIGURE 1.1.** Statistical inference.

chronic bladder infection is twice as great for broccoli eaters when compared to those who do not eat the vegetable. To use statistics in research is happily more than just to compute and report numbers.

The basic process in inferential statistics is to assign probabilities so that we can reach conclusions. The inferences we make are either decisions or estimates about the population. The tool for making inferences is probability (Figure 1.1).

We can illustrate this process by the following example.

---

**Example 1.5. Using Probabilities to Make a Decision**

A sociologist has two large sets of cards, set *A* and set *B*, containing data for her research. The sets each consist of 10,000 cards. Set *A* concerns a group of people, half of whom are women. In set *B*, 80% of the cards are for women. The two files look alike. Unfortunately, the sociologist loses track of which is *A* and which is *B*. She does not want to sort and count the cards, so she decides to use probability to identify the sets. The sociologist selects a set. She draws a card at random from the selected set, notes whether or not it concerns a woman, replaces the card, and repeats this procedure 10 times. She finds that all 10 cards contain data about women. She must now decide between two possible conclusions:

1. This is set *B*.
2. This is set *A*, but an unlikely sample of cards has been chosen.

In order to decide in favor of one of these conclusions, she computes the probabilities of obtaining 10 cards all for females:

$$P(10 \text{ females}) = P(\text{first is female}) \times P(\text{second is female}) \times \cdots \times P(\text{tenth is female})$$

The multiplication rule is used because each choice is independent of the others. For the set *A*, the probability of selecting 10 cards for females is $(0.50)^{10} = 0.00098$ (rounded to two significant digits). For set *B*, the probability of 10 cards for females is $(0.80)^{10} = 0.11$ (again rounded to two significant digits). Since the probability of all 10 of the cards being for women if the set is $B$ is about 100 times the probability if the set is $A$, she decides that the set is $B$, that is, she decides in favor of the conclusion with the higher probability.

---

When we use a strategy based on probability, we are not guaranteed success every time. However, if we repeat the strategy, we will be correct more often than mistaken. In the above example, the sociologist could make the wrong decision because 10 cards chosen at random from set $A$ could all be cards for women. In fact, in repeated experiments using set $A$, 10 cards for females will appear approximately 0.098% of the time, that is, almost once in every thousand 10-card samples.

The example of the files is artificial and oversimplified. In real life, we use statistical methods to reach conclusions about some significant aspect of research in the natural, physical, or social sciences. Statistical procedures do not furnish us with proofs, as do many mathematical techniques. Rather, statistical procedures establish probability bases on which we can accept or reject certain hypotheses.

---

**Example 1.6. Using Probability to Reach a Conclusion in Science**

A real example of the use of statistics in science is the analysis of the effectiveness of Salk's polio vaccine.

A great deal of work had to be done prior to the actual experiment and the statistical analysis. Dr. Jonas Salk first had to gather enough preliminary information and experience in his field to know which of the three polio viruses to use. He had to solve the problem of how to culture that virus. He also had to determine how long to treat the virus with formaldehyde so that it would die but retain its protein shell in the same form as the live virus; the shell could then act as an antigen to stimulate the human body to develop antibodies. At this point, Dr. Salk could conjecture that the dead virus might be used as a vaccine to give patients immunity to paralytic polio.

Finally, Dr. Salk had to decide on the type of experiment that would adequately test his conjecture. He decided on a *double-blind* experiment in which neither patient nor doctor knew whether the patient received the vaccine or a saline solution. The patients receiving the saline solution would form the *control group*, the standard for comparison. Only after all these preliminary steps could the experiment be carried out.

When Dr. Salk speculated that patients inoculated with the dead virus would be immune to paralytic polio, he was formulating the *experimental hypothesis*: the expected outcome if the experimenter's speculation is true. Dr. Salk wanted to use statistics to make a decision about this experimental hypothesis. The decision was to be made solely on the basis of probability. He made the decision in an indirect way; instead of considering the experimental hypothesis itself, he considered a statistical hypothesis called the *null hypothesis*—the expected outcome if the vaccine is ineffective and only chance differences are observed between the two sample groups, the inoculated group and the control group. The null hypothesis is often called the hypothesis of no difference, and it is symbolized $H_0$. In Dr. Salk's experiment, the null hypothesis is that the incidence of paralytic polio in the general population will be the same whether it receives the proposed vaccine or the saline solution. In symbols[†]

$$H_0\colon \pi_{\mathrm{I}} = \pi_{\mathrm{C}}$$

[†]The use of the symbol $\pi$ has nothing to do with the geometry of circles or the irrational number 3.1416 . . . .

in which $\pi_I$ is the proportion of cases of paralytic polio in the general population if it were inoculated with the vaccine and $\pi_C$ is the proportion of cases if it received the saline solution. If the null hypothesis is true, then the two sample groups in the experiment should be alike except for chance differences of exposure and contraction of the disease.

The experimental results were as follows:

| | Proportion with Paralytic Polio | Number in Study |
|---|---|---|
| Inoculated Group | 0.0001603 | 200,745 |
| Control Group | 0.0005703 | 201,229 |

The incidence of paralytic polio in the control group was almost four times higher than in the inoculated group, or in other words the odds ratio was $0.0005703/0.0001603 = 3.56$.

Dr. Salk then found the probability that these experimental results or more extreme ones could have happened with a true null hypothesis. The probability that $\pi_I = \pi_C$ and the difference between the two experimental groups was caused by chance was less than 1 in 10,000,000, so Salk rejected the null hypothesis and decided that he had found an effective vaccine for the general public.[†]

---

Usually when we experiment, the results are not as conclusive as the result obtained by Dr. Salk. The probabilities will always fall between 0 and 1, and we have to establish a level below which we reject the null hypothesis and above which we accept the null hypothesis. If the probability associated with the null hypothesis is small, we reject the null hypothesis and accept an alternative hypothesis (usually the experimental hypothesis). When the probability associated with the null hypothesis is large, we accept the null hypothesis. This is one of the basic procedures of statistical methods—to ask: What is the probability that we would get these experimental results (or more extreme ones) with a true null hypothesis?

Since the experiment has already taken place, it may seem after the fact to ask for the probability that only chance caused the difference between the observed results and the null hypothesis. Actually, when we calculate the probability associated with the null hypothesis, we are asking: If this experiment were performed over and over, what is the probability that chance will produce experimental results as different as are these results from what is expected on the basis of the null hypothesis?

We should also note that Salk was interested not only in the samples of 401,974 people who took part in the study; he was also interested in *all* people, then and in the future, who could receive the vaccine. He wanted to make an inference to the entire population from the portion of the population that he was able to observe. This is called the *target population*, the population about which the inference is intended.

Sometimes in science the inference we should like to make is not in the form of a decision about a hypothesis; but rather it consists of an estimate. For example, perhaps we want to estimate the proportion of adult Americans who approve of the way in which the president is handling the economy, and we want to include some statement about the amount of error possibly related to this estimate. Estimation of this type is another kind of inference, and it also depends on probability. For simplicity, we focus on tests of hypotheses in this

[†]This probability is found using a chi-square test (see Section 5.3).

introductory chapter. The first example of inference in the form of estimation is discussed in Chapter 3.

## EXERCISES

**1.1.1.** A trial mailing is made to advertise a new science dictionary. The trial mailing list is made up of random samples of current mailing lists of several popular magazines. The number of advertisements mailed and the number of people who ordered the dictionary are as follows:

| | Magazine | | | | |
|---|---|---|---|---|---|
| | *A* | *B* | *C* | *D* | *E* |
| Mailed: | 900 | 810 | 1100 | 890 | 950 |
| Ordered: | 18 | 15 | 10 | 30 | 45 |

**a.** Estimate the probability and the odds that a subscriber to each of the magazines will buy the dictionary.

**b.** Make a decision about the mailing list that will probably produce the highest percentage of sales if the entire list is used.

**1.1.2.** In Examples 1.5 and 1.6, probability was used to make decisions and odds ratios could have been used to further support the decisions. To do so:

**a.** For the data in Example 1.5, compute the odds ratio for the two sets of cards.

**b.** For the data in Example 1.6, compute the odds ratio of getting polio for those vaccinated as opposed to those not vaccinated.

**1.1.3.** If 60% of the population of the United States need to have their vision corrected, we say that the probability that an individual chosen at random from the population needs vision correction is $P(C) = 0.60$.

**a.** Estimate the probability that an individual chosen at random does not need vision correction. *Hint:* Use the complement of a probability.

**b.** If 3 people are chosen at random from the population, what is the probability that all 3 need correction, $P(CCC)$? *Hint:* Use the multiplication law of probability for independent events.

**c.** If 3 people are chosen at random from the population, what is the probability that the second person does not need correction but the first and the third do, $P(CNC)$?

**d.** If 3 people are chosen at random from the population, what is the probability that 1 out of the 3 needs correction, $P(CNN$ or $NCN$ or $NNC)$? *Hint:* Use the addition law of probability for mutually exclusive events.

**e.** Assuming no association between vision and gender, what is the probability that a randomly chosen female needs vision correction, $P(C|F)$?

**1.1.4.** On a single roll of 2 dice (think of one green and the other red to keep track of all outcomes) in the game of craps, find the probabilities for:

**a.** A sum of 6, $P(y = 6)$

**b.** A sum of 8, $P(y = 8)$

**c.** A win on the first roll; that is, a sum of 7 or 11, $P(y = 7 \text{ or } 11)$

**d.** A loss on the first roll; that is, a sum of 2, 3, or 12, $P(y = 2, 3, \text{ or } 12)$

**1.1.5.** The dice game about which Pascal and de Fermat were asked consisted in throwing a pair of dice 24 times. The problem was to decide whether or not to bet even money on the occurrence of at least one "double 6" during the 24 throws of a pair of dice. Because it is easier to solve this problem by finding the complement, take the following steps:

**a.** What is the probability of not a double 6 on a roll, $P(E) = P(y \neq 12)$?

**b.** What is the probability that $y = 12$ on all 24 rolls, $P(E_1E_2, \ldots, E_{24})$?

**c.** What is the probability of at least one double 6?

**d.** What are the odds of a win in this game?

**1.1.6.** Sir Francis Galton (1822–1911) was educated as a physician but had the time, money, and inclination for research on whatever interested him, and almost everything did. Though not the first to notice that he could find no two people with the same fingerprints, he was the first to develop a system for categorizing fingerprints and to persuade Scotland Yard to use fingerprints in criminal investigation. He supported his argument with fingerprints of friends and volunteers solicited through the newspapers, and for all comparisons $P(\text{fingerprints match}) = 0$. To compute the number of events associated with Galton's data:

**a.** Suppose fingerprints on only 10 individuals are involved.

- **i.** How many comparisons between individuals can be made? *Hint:* Fingerprints of the first individual can be compared to those of the other 9. However, for the second individual there are only 8 additional comparisons because his fingerprints have already been compared to the first.
- **ii.** How many comparisons between fingers can be made? Assume these are between corresponding fingers of both individuals in a comparison, right thumb of one versus right thumb of the other, and so on.

**b.** Suppose fingerprints are available on 11 individuals rather than 10. Use the results already obtained to simplify computations in finding the number of comparisons among people and among fingers.

## 1.2. THE SCIENTIFIC METHOD

The natural, physical, and social scientists who use statistical methods to reach conclusions all approach their problems by the same general procedure, the *scientific method*. The steps involved in the scientific method are:

1. State the problem.
2. Formulate the hypothesis.
3. Design the experiment or survey.
4. Make observations.
5. Interpret the data.
6. Draw conclusions.

We use statistics mainly in step 5, "interpret the data." In an indirect way we also use statistics in steps 2 and 3, since the formulation of the hypothesis and the design of the experiment or survey must take into consideration the type of statistical procedure to be used in analyzing the data.

The main purpose of this book is to examine step 5. We frequently discuss the other steps, however, because an understanding of the total procedure is important. A statistical analysis may be flawless, but it is not valid if data are gathered incorrectly. A statistical analysis may not even be possible if a question is formulated in such a way that a statistical hypothesis cannot be tested. Considering all of the steps also helps those who study statistical methods before they have had much practical experience in using the scientific method. A full discussion of the scientific method is outside the scope of this book, but in this section we make some comments on the five steps.

***Step 1. State the Problem***. Sometimes, when we read reports of research, we get the impression that research is a very orderly analytic process. Nothing could be further from the truth. A great deal of hidden work and also a tremendous amount of intuition are involved before a solvable problem can even be stated. Technical information and experience are indispensable before anyone can hope to formulate a reasonable problem, but they are not sufficient. The mediocre scientist and the outstanding scientist may be equally familiar with their field; the difference between them is the intuitive insight and skill that the outstanding scientist has in identifying relevant problems that he or she can reasonably hope to solve.

One simple technique for getting a problem in focus is to formulate a clear and explicit statement of the problem and put the statement in writing. This may seem like an unnecessary instruction for a research scientist; however, it is frequently not followed. The consequence is a vagueness and lack of focus that make it almost impossible to proceed. It leads to the collection of unnecessary information or the failure to collect essential information. Sometimes the original question is even lost as the researcher gets involved in the details of the experiment.

***Step 2. Formulate the Hypothesis***. The "hypothesis" in this step is the experimental hypothesis, the expected outcome if the experimenter's speculations are true. The experimental hypothesis must be stated in a precise way so that an experiment can be carried out that will lead to a decision about the hypothesis. A good experimental hypothesis is comprehensive enough to explain a phenomenon and predict unknown facts and yet is stated in a simple way. Classic examples of good experimental hypotheses are Mendel's laws, which can be used to explain hereditary characteristics (such as the color of flowers) and to predict what form the characteristics will take in the future.

Although the null hypothesis is not used in a formal way until the data are being interpreted, it is appropriate to formulate the null hypothesis at this time in order to verify that the experimental hypothesis is stated in such a way that it can be tested by statistical techniques.

Several experimental hypotheses may be connected with a single problem. Once these hypotheses are formulated in a satisfactory way, the investigator should do a literature search to see whether the problem has already been solved, whether or not there is hope of solving it, and whether or not the answer will make a worthwhile contribution to the field.

***Step 3. Design the Experiment or Survey***. Included in this step are several decisions. What *treatments* or conditions should be placed on the objects or subjects of the investigation in order to test the hypothesis? What are the *variables* of interest, that is, what variables should be measured? How will this be done? With how much precision? Each of these decisions is complex and requires experience and insight into the particular area of investigation.

Another group of decisions involves the choice of the *sample*, that portion of the population of interest that will be used in the study. The investigator usually tries to utilize samples that are:

(a) Random
(b) Representative
(c) Sufficiently large

In order to make a decision based on probability, it is necessary that the sample be *random*. Random samples make it possible to determine the probabilities associated with the study. A sample is random if it is just as likely that it will be picked from the population of interest as any other sample of that size. Strictly speaking, statistical inference is not possible unless random samples are used. (Specific methods for achieving random samples are discussed in Section 2.2.)

Random, however, does not mean haphazard. Haphazard processes often have hidden factors that influence the outcome. For example, one scientist using guinea pigs thought that time could be saved in choosing a treatment group and a control group by drawing the treatment group of animals from a box without looking. The scientist drew out half of the guinea pigs for testing and reserved the rest for the control group. It was noticed, however, that most of the animals in the treatment group were larger than those in the control group. For some reason, perhaps because they were larger, or slower, the heavier guinea pigs were drawn first. Instead of this haphazard selection, the experimenter could have recorded the animals' ear-tattoo numbers on plastic disks and drawn the disks at random from a box.

Unfortunately, in many fields of investigation random sampling is not possible, for example, meteorology, some medical research, and certain areas of economics. Random samples are the ideal, but sometimes only nonrandom data are available. In these cases the investigator may decide to proceed with statistical inference, realizing, of course, that it is somewhat risky. Any final report of such a study should include a statement of the author's awareness that the requirement of randomness for inference has not been met.

The second condition that an investigator often seeks in a sample is that it be *representative*. Usually we do not know how to find truly representative samples. Even when we think we can find them, we are often governed by a subconscious bias.

A classic example of a subconscious bias occurred at a Midwestern agricultural station in the early days of statistics. Agronomists were trying to predict the yield of a certain crop in a field. To make their prediction, they chose several 6-ft $\times$ 6-ft sections of the field which they felt were representative of the crop. They harvested those sections, calculated the arithmetic average of the yields, then multiplied this average by the number of 36-ft$^2$ sections in the field to estimate the total yield. A statistician assigned to the station suggested that instead they should have picked random sections. After harvesting several random sections, a second average was calculated and used to predict the total yield. At harvest time, the actual yield of the field was closer to the yield predicted by the statistician. The agronomists had predicted a much larger yield, probably because they chose sections that looked like an ideal crop. An entire field, of course, is not ideal. The unconscious bias of the agronomists prevented them from picking a representative sample. Such unconscious bias cannot occur when experimental units are chosen at random.

Although representativeness is an intuitively desirable property, in practice it is usually an impossible one to meet. How can a sample of 30 possibly contain all the properties of a population of 2000 individuals? The 2000 certainly have more characteristics than can possibly be proportionately reflected in 30 individuals. So although representativeness seems necessary for proper reasoning from the sample to the population, statisticians do not rely on representative samples—rather, they rely on random samples. (Large random samples will very likely be representative). If we do manage to deliberately construct a sample that is representative but is not random, we will be unable to compute probabilities related to the sample and, strictly speaking, we will be unable to do statistical inference.

It is also necessary that samples be *sufficiently large*. No one would question the necessity of repetition in an experiment or survey. We all know the danger of generalizing from a single observation. Sufficiently large, however, does not mean massive repetition. When we use statistics, we are trying to get information from relatively small samples. Determining a reasonable sample size for an investigation is often difficult. The size depends upon the magnitude of the difference we are trying to detect, the variability of the variable of interest, the type of statistical procedure we are using, the seriousness of the errors we might make, and the cost involved in sampling. (We make further remarks on sample size as we discuss various procedures throughout this text.)

***Step 4. Make Observations***. Once the procedure for the investigation has been decided upon, the researcher must see that it is carried out in a rigorous manner. The study should be free from all errors except random measurement errors, that is, slight variations that are due to the limitations of the measuring instrument.

Care should be taken to avoid *bias*. Bias is a tendency for a measurement on a variable to be affected by an external factor. For example, bias could occur from an instrument out of calibration, an interviewer who influences the answers of a respondent, or a judge who sees the scores given by other judges. Equipment should not be changed in the middle of an experiment, and judges should not be changed halfway through an evaluation.

The data should be examined for unusual values, *outliers*, which do not seem to be consistent with the rest of the observations. Each outlier should be checked to see whether or not it is due to a recording error. If it is an error, it should be corrected. If it cannot be corrected, it should be discarded. If an outlier is not an error, it should be given special attention when the data are analyzed. For further discussion, see Barnett and Lewis (2002).

Finally, the investigator should keep a complete, legible record of the results of the investigation. All original data should be kept until the analysis is completed and the final report written. Summaries of the data are often not sufficient for a proper statistical analysis.

***Step 5. Interpret the Data***. The general statistical procedure was illustrated in Example 1.6, in which the Salk vaccine experiment was discussed. To interpret the data, we set up the null hypothesis and then decide whether the experimental results are a rare outcome if the null hypothesis is true. That is, we decide whether the difference between the experimental outcome and the null hypothesis is due to more than chance; if so, this indicates that the null hypothesis should be rejected.

If the results of the experiment are unlikely when the null hypothesis is true, we reject the null hypothesis; if they are expected, we accept the null hypothesis. We must remember, however, that statistics does not prove anything. Even Dr. Salk's result, with a probability of less than 1 in 10,000,000 that chance was causing the difference between the experimental outcome and the null hypothesis, does not prove that the null hypothesis is false. An extremely small probability, however, does make the scientist believe that the difference is not due to chance alone and that some additional mechanism is operating.

Two slightly different approaches are used to evaluate the null hypothesis. In practice, they are often intermingled. Some researchers compute the probability that the experimental results, or more extreme values, could occur if the null hypothesis is true; then they use that probability to make a judgment about the null hypothesis. In research articles this is often reported as the *observed significance level*, or the *significance level*, or the *P value*. If the $P$ value is large, they conclude that the data are consistent with the null hypothesis. If the $P$ value is small, then either the null hypothesis is false or the null hypothesis is true and a rare event has occurred. (This was the approach used in the Salk vaccine example.)

Other researchers prefer a second, more decisive approach. Before the experiment they decide on a *rejection level*, the probability of an unlikely event (sometimes this is also called the *significance level*). An experimental outcome, or a more extreme one, that has a probability below this level is considered to be evidence that the null hypothesis is false. Some research articles are written with this approach. It has the advantage that only a limited number of probability tables are necessary. Without a computer, it is often difficult to determine the exact $P$ value needed for the first approach. For this reason the second approach became popular in the early days of statistics. It is still frequently used.

The sequence in this second procedure is:

(a) Assume $H_0$ is true and determine the probability $P$ that the experimental outcome or a more extreme one would occur.
(b) Compare the probability to a preset rejection level symbolized by $\alpha$ (the Greek letter alpha).
(c) If $P \leq \alpha$, reject $H_0$. If $P > \alpha$, accept $H_0$.

If $P > \alpha$, we say, “Accept the null hypothesis.” Some statisticians prefer not to use that expression, since in the absence of evidence to reject the null hypothesis, they choose simply to withhold judgment about it. This group would say, “The null hypothesis may be true” or “There is no evidence that the null hypothesis is false.”

If the probability associated with the null hypothesis is very close to $\alpha$, more extensive testing may be desired. Notice that this is a blend of the two approaches.

An example of the total procedure follows.

---

**Example 1.7. Using a Statistical Procedure to Interpret Data**

A manufacturer of baby food gives samples of two types of baby cereal, $A$ and $B$, to a random sample of four mothers. Type $A$ is the manufacturer’s brand, type $B$ a competitor’s. The mothers are asked to report which type they prefer. The manufacturer wants to detect any preference for their cereal if it exists.

The null hypothesis, or the hypothesis of no difference, is $H_0$: $\pi = 1/2$, in which $\pi$ is the proportion of mothers in the general population who prefer type $A$. The experimental hypothesis, which often corresponds to a second statistical hypothesis called the *alternative hypothesis*, is that there is a preference for cereal $A$, $H_a$: $\pi > 1/2$.

Suppose that four mothers are asked to choose between the two cereals. If there is no preference, the following 16 outcomes are possible with equal probability:

| | | | |
|---|---|---|---|
| *AAAA* | *AAAB* | *ABBA* | *BBAB* |
| *BAAA* | *BBAA* | *ABAB* | *BABB* |
| *ABAA* | *BABA* | *AABB* | *ABBB* |
| *AABA* | *BAAB* | *BBBA* | *BBBB* |

The manufacturer feels that only 1 of these 16 cases, *AAAA*, is very different from what would be expected to occur under random sampling, when the null hypothesis of no preference is true. Since the unusual case would appear only 1 time out of 16 times when the null hypothesis is true, $\alpha$ (the rejection level) is set equal to $1/16 = 0.0625$.

If the outcome of the experiment is in fact four choices of type *A*, then $P = P(AAAA) = 1/16$, and the manufacturer can say that the results are in the region of rejection, or the results are significant, and the null hypothesis is rejected. If the outcome is three choices of type *A*, however, then $P = P(3 \text{ or more } A\text{'s}) = P(AAAB \text{ or } AABA \text{ or } ABAA \text{ or } BAAA \text{ or } AAAA) = 5/16 > 1/16$, and he does not reject the null hypothesis. (Notice that *P* is the probability of this type of outcome or a more extreme one in the direction of the alternative hypothesis, so *AAAA* must be included.)

---

The way in which we set the rejection level $\alpha$ depends on the field of research, on the seriousness of an error, on cost, and to a great degree on tradition. In the example above, the sample size is 4, so an $\alpha$ smaller than $1/16$ is impossible. Later (in Section 3.2), we discuss using the seriousness of errors to determine a reasonable $\alpha$. If the possible errors are not serious and cost is not a consideration, traditional values are often used.

Experimental statistics began about 1920 and was not used much until 1940, but it is already tradition bound. In the early part of the twentieth century Karl Pearson had his students at University College, London, compute tables of probabilities for reasonably rare events. Now computers are programmed to produce these tables, but the traditional levels used by Pearson persist for the most part. Tables are usually calculated for $\alpha$ equal to 0.10, 0.05, and 0.01. Many times there is no justification for the use of one of these values except tradition and the availability of tables. If an $\alpha$ close to but less than or equal to 0.05 were desired in the example above, a sample size of at least 5 would be necessary, then $\alpha = 1/32 = 0.03125$ if the only extreme case is *AAAAA*.

Step 6. Draw Conclusions. If the procedure just outlined is followed, then our decisions will be based solely on probability and will be consistent with the data from the experiment. If our experimental results are not unusual for the null hypothesis, $P > \alpha$, then the null hypothesis seems to be right and we should not reject it. If they are unusual, $P \leq \alpha$, then the null hypothesis seems to be wrong and we should reject it. We repeat that our decision could be incorrect, since there is a small probability $\alpha$ that we will reject a null hypothesis when in fact that null hypothesis is true; there is also a possibility that a false null hypothesis will be accepted. (These possible errors are discussed in Section 3.2.)

In some instances, the conclusion of the study and the statistical decision about the null hypothesis are the same. The conclusion merely states the statistical decision in specific terms. In many situations, the conclusion goes further than the statistical decision. For example, suppose that an orthodontist makes a study of malocclusion due to crowding of the adult lower front teeth. The orthodontist hypothesizes that the incidence is as common in males as in females, $H_0$: $\pi_M = \pi_F$. (Note that in this example the experimental hypothesis coincides with the null hypothesis.) In the data gathered, however, there is a preponderance of males and $P \leq \alpha$. The statistical decision is to reject the null hypothesis, but this is not the final statement. Having rejected the null hypothesis, the orthodontist concludes the report by stating that this condition occurs more frequently in males than in females and advises family dentists of the need to watch more closely for tendencies of this condition in boys than in girls.

## EXERCISES

**1.2.1.** Put the example of the cereals in the framework of the scientific method, elaborating on each of the six steps.

**1.2.2.** State a null and alternative hypotheses for the example of the file cards in Section 1.1, Example 1.5.

**1.2.3.** In the Salk experiment described in Example 1.6 of Section 1.1:

**a.** Why should Salk not be content just to reject the null hypothesis?

**b.** What conclusion could be drawn from the experiment?

**1.2.4.** Two college roommates decide to perform an experiment in extrasensory perception (ESP). Each produces a snapshot of his home-town girl friend, and one snapshot is placed in each of two identical brown envelopes. One of the roommates leaves the room and the other places the two envelopes side by side on the desk. The first roommate returns to the room and tries to pick the envelope that contains his girl friend's picture. The experiment is repeated 10 times. If the one who places the envelopes on the desk tosses a coin to decide which picture will go to the left and which to the right, the probabilities for correct decisions are listed below.

| Number of Correct Decisions | Probability | Number of Correct Decisions | Probability |
|---|---|---|---|
| 0 | 1/1024 | 6 | 210/1024 |
| 1 | 10/1024 | 7 | 120/1024 |
| 2 | 45/1024 | 8 | 45/1024 |
| 3 | 120/1024 | 9 | 10/1024 |
| 4 | 210/1024 | 10 | 1/1024 |
| 5 | 252/1024 | | |

**a.** State the null hypothesis based on chance as the determining factor in a correct decision. (Make the statement in words and symbols.)

**b.** State an alternative hypothesis based on the power of love.

**c.** If $\alpha$ is set as near 0.05 as possible, what is the region of rejection, that is, what numbers of correct decisions would provide evidence for ESP?

**d.** What is the *region of acceptance*, that is, those numbers of correct decisions that would not provide evidence of ESP?

**e.** Suppose the first roommate is able to pick the envelope containing his girl friend's picture 10 times out of 10; which of the following statements are true?

**i.** The null hypothesis should be rejected.

**ii.** He has demonstrated ESP.

**iii.** Chance is not likely to produce such a result.

**iv.** Love is more powerful than chance.

**v.** There is sufficient evidence to suspect that something other than chance was guiding his selections.

**vi.** With his luck he should raise some money and go to Las Vegas.

**1.2.5.** The mortality rate of a certain disease is 50% during the first year after diagnosis. The chance probabilities for the number of deaths within a year from a group of six persons with the disease are:

| Number of deaths: | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Probability: | 1/64 | 6/64 | 15/64 | 20/64 | 15/64 | 6/64 | 1/64 |

A new drug has been found that is helpful in cases of this disease, and it is hoped that it will lower the death rate. The drug is given to 6 persons who have been diagnosed as having the disease. After a year, a statistical test is performed on the outcome in order to make a decision about the effectiveness of the drug.

**a.** What is the null hypothesis, in words and symbols?

**b.** What is the alternative hypothesis, based on the prior evidence that the drug is of some help?

**c.** What is the region of rejection if $\alpha$ is set as close to 0.10 as possible?

**d.** What is the region of acceptance?

**e.** Suppose that 4 of the 6 persons die within one year. What decision should be made about the drug?

**1.2.6.** A company produces a new kind of decaffeinated coffee which is thought to have a taste superior to the three currently most popular brands. In a preliminary random sample, 20 consumers are presented with all 4 kinds of coffee (in unmarked containers and in random order), and they are asked to report which one tastes best. If all 4 taste equally good, there is a 1-in-4 chance that a consumer will report that the new product tastes best. If there is no difference, the probabilities for various numbers of consumers indicating by chance that the new product is best are:

| Number picking new product: | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Probability: | 0.003 | 0.021 | 0.067 | 0.134 | 0.190 |
| Number picking new product: | 5 | 6 | 7 | 8 | 9 |
| Probability: | 0.202 | 0.169 | 0.112 | 0.061 | 0.027 |
| Number picking new product: | 10 | 11 | 12 | 13–20 | |
| Probability: | 0.010 | 0.003 | 0.001 | <0.001 | |

**a.** State the null and alternative hypotheses, in words and symbols.

**b.** If $\alpha$ is set as near 0.05 as possible, what is the region of rejection? What is the region of acceptance?

**c.** Suppose that 6 of the 20 consumers indicate that they prefer the new product. Which of the following statements is correct?

**i.** The null hypothesis should be rejected.

**ii.** The new product has a superior taste.

**iii.** The new product is probably inferior because fewer than half of the people selected it.

**iv.** There is insufficient evidence to support the claim that the new product has a superior taste.

## 1.3. EXPERIMENTAL DATA AND SURVEY DATA

An *experiment* involves the collection of measurements or observations about populations that are treated or controlled by the experimenter. A *survey*, in contrast to an experiment, is an examination of a system in operation in which the investigator does not have an opportunity to assign different conditions to the objects of the study. Both of these methods of data collection may be the subject of statistical analysis; however, in the case of surveys some cautions are in order.

We might use a survey to compare two countries with different types of economic systems. If there is a significant difference in some economic measure, such as per-capita income, it does not mean that the economic system of one country is superior to the other. The survey takes conditions as they are and cannot control other variables that may affect the economic measure, such as comparative richness of natural resources, population health, or level of literacy. All that can be concluded is that at this particular time a significant difference exists in the economic measure. Unfortunately, surveys of this type are frequently misinterpreted.

A similar mistake could have been made in a survey of the life expectancy of men and women. The life expectancy was found to be 74.1 years for men and 79.5 years for women. Without control for risk factors—smoking, drinking, physical inactivity, stressful occupation, obesity, poor sleeping patterns, and poor life satisfaction—these results would be of little value. Fortunately, the investigators gathered information on these factors and found that women have more high-risk characteristics than men but still live longer. Because this was a carefully planned survey, the investigators were able to conclude that women biologically have greater longevity.

Surveys in general do not give answers that are as clear-cut as those of experiments. If an experiment is possible, it is preferred. For example, in order to determine which of two methods of teaching reading is more effective, we might conduct a survey of two schools that are each using a different one of the methods. But the results would be more reliable if we could conduct an experiment and set up two balanced groups within one school, teaching each group by a different method.

From this brief discussion it should not be inferred that surveys are not trustworthy. Most of the data presented as evidence for an association between heavy smoking and lung cancer come from surveys. Surveys of voter preference cause certain people to seek the presidency and others to decide not to enter the campaign. Quantitative research in many areas of social, biological, and behavioral science would be impossible without surveys. However, in surveys we must be alert to the possibility that our measurements may be affected by variables that are not of primary concern. Since we do not have as much control over these variables as we have in an experiment, we should record all concomitant information of pertinence for each observation. We can then study the effects of these other variables on the variable of interest and possibly adjust for their effects.

## EXERCISES

**1.3.1.** In each of the research situations described below, determine whether the researcher is conducting an experiment or a survey.

- **a.** Traps are set out in a grain field to determine whether rabbits or raccoons are the more frequently found pests.
- **b.** A graduate student in English literature uses random 500-word passages from the writings of Shakespeare and Marlowe to determine which author uses the conditional tense more frequently.
- **c.** A random sample of hens is divided into 2 groups at random. The first group is given minute quantities of an insecticide containing an organic phosphorus compound; the second group acts as a control group. The average difference in eggshell thickness between the 2 groups is then determined.
- **d.** To determine whether honeybees have a color preference in flowers, an apiarist mixes a sugar-and-water solution and puts equal amounts in 2 equal-sized sets of vials of different colors. Bees are introduced into a cage containing the vials, and the frequency with which bees visit vials of each color is recorded.

**1.3.2.** In each of the following surveys, what besides the mechanism under study could have contributed to the result?

- **a.** An estimation of per-capita wealth for a city is made from a random sample of people listed in the city's telephone directory.
- **b.** Political preference is determined by an interviewer taking a random sample of Monday morning bank customers.
- **c.** The average length of fish in a lake is estimated by:
  - **i.** The average length of fish caught, reported by anglers
  - **ii.** The average length of dead fish found floating in the water
- **d.** The average number of words in the working vocabulary of first-grade children in a given county is estimated by a vocabulary test given to a random sample of first-grade children in the largest school in the country.
- **e.** The proportion of people who can distinguish between two similar tones is estimated on the basis of a test given to a random sample of university students in a music appreciation class.

**1.3.3.** *Time* magazine once reported that El Paso's water was heavily laced with lithium, a tranquilizing chemical, whereas Dallas had a low lithium level. *Time* also reported that FBI statistics showed that El Paso had 2889 known crimes per 100,000 population and Dallas had 5970 known crimes per 100,000 population. The article reported that a University of Texas biochemist felt that the reason for the lower crime rate in El Paso lay in El Paso's water. Comment on the biochemist's conjecture.

## 1.4. COMPUTER USAGE

The practice of statistics has been radically changed now that computers and high-quality statistical software are readily available and relatively inexpensive. It is no longer necessary to spend large amounts of time doing the numerous calculations that are part of a statistical analysis. We need only enter the data correctly, choose the appropriate procedure, and then have the computer take care of the computational details.

Because the computer can do so much for us, it might seem that it is now unnecessary to study statistics. Nothing could be further from the truth. Now more than ever the researcher needs a solid understanding of statistical analysis. The computer does not choose the statistical procedure or make the final interpretation of the results; these steps are still in the hands of the investigator.

Statistical software can quickly produce a large variety of analyses on data regardless of whether these analyses correspond to the way in which the data were collected. An inappropriate analysis yields results that are meaningless. Therefore, the researcher must learn the conditions under which it is valid to use the various analyses so that the selection can be made correctly.

The computer program will produce a numerical output. It will not indicate what the numbers mean. The researcher must draw the statistical conclusion and then translate it into the concrete terms of the investigation. Statistical analysis can best be described as a search for evidence. What the evidence means and how much weight to give to it must be decided by the researcher.

In this text we have included some computer output to illustrate how the output could be used to perform some of the analyses that are discussed. Several exercises have computer output to assist the user with analyzing the data. Additional output illustrating nearly all the procedures discussed is available on an Internet website.

Many different comprehensive statistical software packages are available and the outputs are very similar. A researcher familiar with the output of one package will probably find it easy to understand the output of a different package. We have used two particular packages, the SAS system and JMP, for the illustrations in the text. The SAS system was designed originally for batch use on the large mainframe computers of the 1970's. JMP was originally designed for interactive use on the personal computers of the 1980's. SAS made it possible to analyze very large sets of data simply and efficiently. JMP made it easy to visualize smaller sets of data. Because the distinction between large and small is frequently unclear, it is useful to know about both programs.

The computer could be used to do many of the exercises in the text; however, some calculations by the reader are still necessary in order to keep the computer from becoming a magic box. It is easier for the investigator to select the right procedure and to make a proper interpretation if the method of computation is understood.

## REVIEW EXERCISES

Decide whether each of the following statements is true or false. If a statement is false, explain why.

**1.1.** To say that the null hypothesis is rejected does not necessarily mean it is false.

**1.2.** In a practical situation, the null hypothesis, alternative hypothesis, and level of rejection should be specified before the experimentation.

**1.3.** The probability of choosing a random sample of 3 persons in which the first 2 say "yes" and the last person says "no" from a population in which $P(\text{yes}) = 0.7$ is (0.7)(0.7)(0.3).

**1.4.** If the experimental hypothesis is true, chance does not enter into the outcome of the experiment.

**1.5.** The alternative hypothesis is often the experimental hypothesis.

**1.6.** A decision made on the basis of a statistical procedure will always be correct.

**1.7.** The probability of choosing a random sample of 3 persons in which exactly 2 say "yes" from a population with $P(\text{yes}) = 0.6$ is (0.6)(0.6)(0.4).

**1.8.** In the total process of investigating a question, the very first thing a scientist does is state the problem.

**1.9.** A scientist completes an experiment and then forms a hypothesis on the basis of the results of the experiment.

**1.10.** In an experiment, the scientist should always collect as large an amount of data as is humanly possible.

**1.11.** Even a specialist in a field may not be capable of picking a sample that is truly representative, so it is better to choose a random sample.

**1.12.** If in an experiment $P(\text{success}) = 1/3$, then the odds against success are 3 to 1.

**1.13.** One of the main reasons for using random sampling is to find the probability that an experiment could yield a particular outcome by chance if the null hypothesis is true.

**1.14.** The $\alpha$ level in a statistical procedure depends on the field of investigation, the cost, and the seriousness of error; however, traditional levels are often used.

**1.15.** A conclusion reached on the basis of a correctly applied statistical procedure is based solely on probability.

**1.16.** The null hypothesis may be the same as the experimental hypothesis.

**1.17.** The "$\alpha$ level" and the "region of rejection" are two expressions for the same thing.

**1.18.** If a correct statistical procedure is used, it is possible to reject a true null hypothesis.

**1.19.** The probability of rolling two 6's on two dice is $1/6 + 1/6 = 1/3$.

**1.20.** A weakness of many surveys is that there is little control of secondary variables.

## SELECTED READINGS

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, **2**, 123–147.

Barnard, G. A. (1947). The meaning of a significance level. *Biometrika*, **34**, 179–182.

Barnett, V., and T. Lewis (2002). *Outliers in Statistical Data*, 3rd ed. Wiley, New York.

Bennett, J. O., W. L. Briggs, and M. F. Triola (2003). *Statistical Reasoning for Everyday Life*, 2nd ed. Addison-Wesley, New York.

Berkson, J. (1942). Tests of significance considered as evidence. *Journal of the American Statistical Association*, **37**, 325–335.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, **71**, 791–799.

Cox, D. R. (1958). *Planning of Experiments*. Wiley, New York.

Duggan, T. J., and C. W. Dean (1968). Common misinterpretation of significance levels in sociology journals. *American Sociologist*, **3**, 45–46.

Edgington, E. S. (1966). Statistical inference and nonrandom samples. *Psychological Bulletin*, **66**, 485–487.

Edwards, W. (1965). Tactical note on the relation between scientific and statistical hypotheses. *Psychological Bulletin*, **63**, 400–402.

Ehrenberg, A. S. C. (1982). Writing technical papers or reports. *American Statistician*, **36**, 326–329.

Gibbons, J. D., and J. W. Pratt (1975). *P*-values: Interpretation and methodology. *American Statistician*, **29**, 20–25.

Gold, D. (1969). Statistical tests and substantive significance. *American Sociologist*, **4**, 42–46.

Greenberg, B. G. (1951). Why randomize? *Biometrics*, **7**, 309–322.

Johnson, R., and P. Kuby (2000). *Elementary Statistics*, 8th ed. Duxbury Press, Pacific Grove, California.

Labovitz, S. (1968). Criteria for selecting a significance level: A note on the sacredness of .05. *American Sociologist*, **3**, 220–222.

McGinnis, R. (1958). Randomization and inference in sociological research. *American Sociological Review*, **23**, 408–414.

Meier, P. (1990). Polio trial: an early efficient clinical trial. *Statistics in Medicine*, **9**, 13–16.

Plutchik, R. (1974). *Foundations of Experimental Research*, 2nd ed. Harper & Row, New York.

Rosenberg, M. (1968). *The Logic of Survey Analysis*. Basic Books, New York.

Royall, R. M. (1986). The effect of sample size on the meaning of significance tests. *American Statistician*, **40**, 313–315.

Selvin, H. C. (1957). A critique of tests of significance in survey research. *American Sociological Review*, **22**, 519–527.

Stigler, S. M. (1986). *The History of Statistics*. Harvard University Press, Cambridge.