NAVIGATIONS SERIES

NAVIGATING *through* DATA ANALYSIS

GRADES 9–12

Chapter 2
Making Decisions with Categorical Data

In data analysis, a categorical variable *represents a characteristic of the population that can be used to classify the individuals or objects into one and only one group (or category). Examples include gender, political party preference, and country of origin.*

How do you approach a problem or question of interest that has no clear-cut answer? That is, how do you arrive at a conclusion through a decision-making process that involves some degree of uncertainty? The activities in this chapter invite students to explore these questions as they examine an experiment that involves categorical variables. One way to summarize categorical variables in a problem is to count the number of observations in given categories. Analyzing such a problem then usually involves deciding what to do with the counts to determine whether the observation under investigation occurred by chance or resulted from other factors.

This chapter focuses on the problem scenario Discrimination or Not? described at the beginning of chapter 1 (p. 11). As noted in that chapter, statisticians are often asked to look at data when someone believes that discrimination may have taken place. The scenario, which raises a question about possible discrimination, comes from a well-known study conducted by Benson Rosen and Thomas H. Jerdee in 1972 and reported in the *Journal of Applied Psychology* in 1974. Although now thirty years old, the study involves data and statistical thinking that are still relevant today. The activity sheet "Discrimination or Not?" provides the study's scenario, reproduced here for the reader's convenience:

p. 104

> In 1972, 48 male bank supervisors were each randomly assigned a personnel file and asked to judge whether the person represented in the file should be recommended for promotion to a branch-manager job described as "routine" or whether the person's file should be held and other applicants interviewed. The files were all identical except that half of the supervisors had files labeled "male" while the other half had files labeled "female." Of the 48 files reviewed, 35 were recommended for promotion.

Twenty-one of the 35 recommended files were labeled "male," and 14 were labeled "female." Can we conclude that discrimination against women played a role in the supervisors' recommendations? The three activities that accompany chapter 2 allow students to examine this scenario. As they work, they will identify relevant questions to consider about discrimination, create a simulation to investigate the possible presence of discrimination, define the numerical summary or statistic that they will calculate from each simulation, and form and explore the simulated sampling distribution based on the statistics resulting from each simulated sample.

To help your students begin thinking about the scenario, you might give them copies of the activity sheet "Discrimination or Not?" and ask them to brainstorm about the following question: "How could you create a model to determine whether any given distribution of the 35 "recommended" files into "male" and "female" files could have occurred by chance variation?" Students might work in groups to formulate their strategies and then reconvene as a class to discuss how they would use these ideas to analyze the problem. After considering with your students some possible models to evaluate any results from the study, you could turn their attention to the first activity, What Would You Expect?

What Would You Expect?

Goals

Begin a statistical exploration of Discrimination or Not? by completing two-way tables showing hypothetical situations with—

(*a*) discrimination clearly playing no role in the results

(*b*) discrimination against women clearly playing a role in the results

(*c*) the results falling in a "gray" area

Materials

- A copy of the activity pages for each student
- A copy of the activity sheet "Discrimination or Not?" for each student

pp. 104, 105–6

"All students should … formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them."
(NCTM 2000, p. 324)

Discussion

Understanding the problem is essential to solving it. Our scenario involves two categorical variables of interest: *gender* (male, female) and *recommendation* (recommended, not recommended). Gender, the *explanatory variable*, might help explain or provide information about recommendation, the *response variable*. The activity helps students use two-by-two (or two-way) tables such as tables 2.1 and 2.2 to organize the information from the study and to consider different hypothetical scenarios. If the recommendations involved no discrimination, one could reasonably expect that the number of males recommended for promotion and the number of females recommended for promotion would be about the same because the numbers of files for each category of gender were the same. (See table 2.1.) Note that either 17 or 18 males recommended for promotion could be regarded as approximately half of the 35 individuals who were recommended for promotion. The number 17 has been used in the table here.

Table 2.1.
A Hypothetical Situation Involving No Evidence of Discrimination against Females

	Recommended for Promotion	Not Recommended for Promotion	Total
Male	17	7	24
Female	18	6	24
Total	35	13	48

For the table constructed in the activity, gender defines the rows, and recommendation defines the columns. Within the groups of males and females, it is important to consider the percentage in each cell.

By contrast, if the recommendations showed strong evidence of discrimination against the women considered for promotion, the numbers in such a table would look quite different. The activity asks students to complete a table showing the situation if the vast majority of the males were recommended for promotion. Table 2.2 shows the case with *all* the males recommended for promotion. Students may also suggest 22 and 23 as reasonable values for the number of men recommended for promotion if discrimination against women clearly played a role.

Table 2.2.
A Hypothetical Situation Involving Strong Evidence of Discrimination against Females

	Recommended for Promotion	Not Recommended for Promotion	Total
Male	24	0	24
Female	11	13	24
Total	35	13	48

The activity then calls on students to suppose a more complicated hypothetical situation, with the evidence of discrimination against the women falling into a "gray" area. In such a case, the presence of discrimination would not be obvious without further investigation. Students may offer a variety of tables for this case, depending on how they perceive the possible variation around the expected values. Table 2.3 illustrates such a situation with the data from the original study.

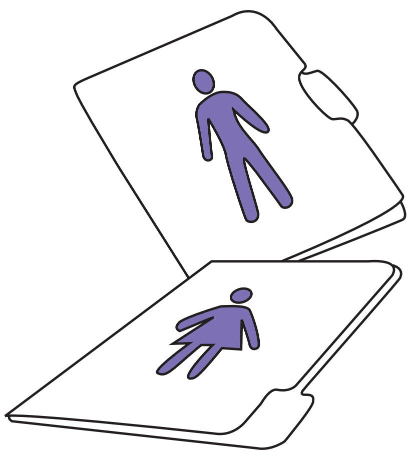


Table 2.3.
Actual Results of the Discrimination Study

	Recommended for Promotion	Not Recommended for Promotion	Total
Male	21	3	24
Female	14	10	24
Total	35	13	48

As the students consider the three scenarios, they should reflect on one of the main questions of interest: How much variability would one expect in the observed counts in a table like table 2.1, which shows the situation when no evidence of discrimination exists? Is it likely that 19 men would be recommended for promotion when you expected 17 or 18? Is it likely that 21 men would be recommended? If this study were repeated many times and discrimination played no part in the recommendations, what counts would be observed? How would the numbers of males and females recommended for promotion fluctuate just by chance variation in the presence of no discrimination?

Simulating the Discrimination Case

In real life, researchers do not usually have the opportunity to repeat a study many times. Simulation, however, is a powerful tool that allows them to model situations and investigate what would happen if they could repeat the study many times. The next activity, Simulating the Case, helps students use simulation to investigate the scenario Discrimination or Not? By allowing students to gather information about the variability that can result by chance when there is no discrimination, the simulation prepares them to use statistical reasoning to answer the real question about discrimination.

Simulating the Case

Goals

- Investigate Discrimination or Not? by—
  (*a*) creating a simulation to model a review of 48 personnel folders
  (*b*) repeating the simulation 20 times to get an idea of the variability to expect from chance when there is no discrimination
- Use the results of the simulation to consider whether the number of males recommended for promotion was the result of chance variation or the result of discrimination against women

Materials

- A copy of the activity pages for each student
- A copy of the activity sheet "Discrimination or Not?" for each student
- A standard deck of 52 playing cards for each student or pair of students

pp. 104, 107–8

Discussion

In Simulating the Case, students use a standard deck of 52 playing cards to create a simple simulation of the situation in "Discrimination or Not?" They remove 2 red cards and 2 black cards, leaving 48 cards divided evenly between black and red. The students can then let the 24 black cards represent the male candidates for promotion and the 24 red cards represent the female candidates, or vice versa.

The structure of the simulation depends on keeping the row and column totals from tables 2.1, 2.2, and 2.3 fixed: That is, (1) after a review of 48 candidates' files, 35 out of the 48 candidates were recommended for promotion, and (2) 24 of the files were labeled "male" and the other 24 were labeled "female."

The object of the activity, then, is to use the 48 cards to simulate the selection process that resulted in the choosing of 35 candidates to recommend for promotion from the 48 personnel files in the study if no discrimination were present. An alternate method involving quick and easy counting is to investigate the 13 candidates who were not recommended for promotion. Dealing 13 cards is simpler and more efficient than dealing 35. Students using this method can then quickly count the cards that represent men who were not recommended for promotion and subtract this number from the total number of male candidates reviewed to get the number of men recommended for promotion. For example, if in a simulation 8 out of the 13 who were not recommended for promotion were men, then the other 16 men would all have been recommended for promotion. In this simulated sample, 16 is the numerical summary, or statistic, that estimates the number of men that one would expect to be recommended for promotion if there were no discrimination.

Simulated counts generated by such a chance process will produce a simulated sampling distribution of counts for males recommended for promotion if there were no discrimination. Students should perform

Students can use the Discrimination or Not applet on the CD-ROM to carry out the simulation.

"All students should … use simulations to explore the variability of sample statistics from a known population and to construct sampling distributions." (NCTM 2000, p. 324)

Recall that a *sampling distribution* is the distribution of the values of a statistic for repeated samples of the same size from a population. For the scenario Discrimination or Not? the number of black cards (representing the number of males recommended for promotion) from each simulation is the *statistic.*

the simulation at least 20 times to get a sense of the center, shape, and spread of this simulated sampling distribution. The observed number of males recommended for promotion (21) can then be compared to the simulated sampling distribution.

You can ask your students to carry out the simulation in one of several ways. As a homework activity, students can independently create their own sets of 20 trials of the simulation—that is, they can deal out samples of 13 cards 20 times, being sure to shuffle the deck thoroughly between deals. Then each student will have a simulation of the sampling distribution of sample counts of males recommended for promotion to analyze. Alternatively, students might conduct only one or two trials independently before the whole class works together to pool results and analyze them. Another approach is to assign the activity for students to complete in pairs.

Direct your students' attention again to the tables in the activity What Would You Expect? and point out that only one of the observed counts in the table cells can vary freely, because the row and column totals are fixed. For example, once the number of males recommended for promotion has been determined, the other three cell values must be counts that add to the fixed row and column totals. Thus, data analysis can be based on just one of the cell categories. Here the number of males recommended for promotion is the category used.

Chapter 1 introduced ideas about sampling, sampling distributions, and variability and explained the concept of a *statistic*. The statistic being considered here is the number of men recommended for promotion. The observed value of this statistic from the actual study is 21. In the activity Simulating the Case, students use simulation to create a simulated sampling distribution of the statistics obtained from each simulation. The overriding goal of the activity is to provide information to answer the following question: "Is the *observed* number of males recommended for promotion greater than the number that would be *expected* as a result of chance variation?"

The next activity, Analyzing Simulation Results, presents three sample student simulations that teachers can use as an exercise for students to work on individually or together as a whole class.

Analyzing Simulation Results

Goals

- Continue to investigate the scenario Discrimination or Not? by comparing and analyzing the shape, center, and spread of the simulated sampling distributions based on three sets of student simulations
- Use the simulated sampling distributions to consider how infrequent an event must be for it to be regarded as “rare”

Materials

- A copy of the activity pages for each student
- A copy of the activity sheet “Discrimination or Not?” for each student

Discussion

Figures 2.1, 2.2, and 2.3 present the results of three sets of student simulations that can help answer the question about discrimination raised by the scenario Discrimination or Not?

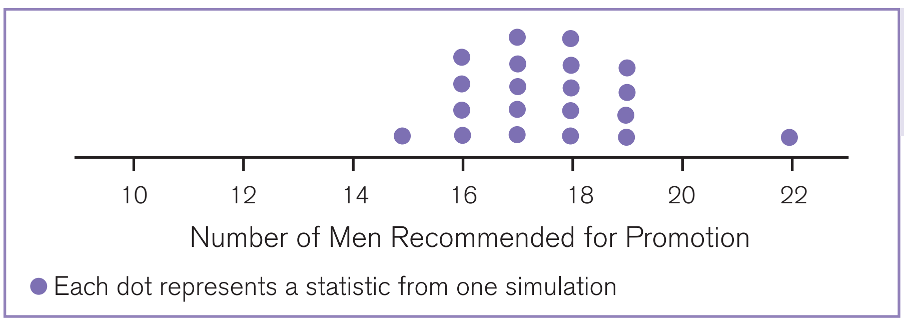


Fig. **2.1.**

The number of men recommended for promotion in 20 simulations

Descriptions of any distribution should focus on the shape, center, and spread or variability of the distribution. Students can see that the shape of the simulated sampling distribution in figure 2.1 is approximately symmetrical, with one unusual observation at 22 males. The value at 22 is a potential *outlier*—that is, an observation that does not fall into the overall pattern of the other observations in the distribution. Half of the counts are higher than the median of 17.5 males. The mean, or average, of the simulated sampling distribution appears to be around

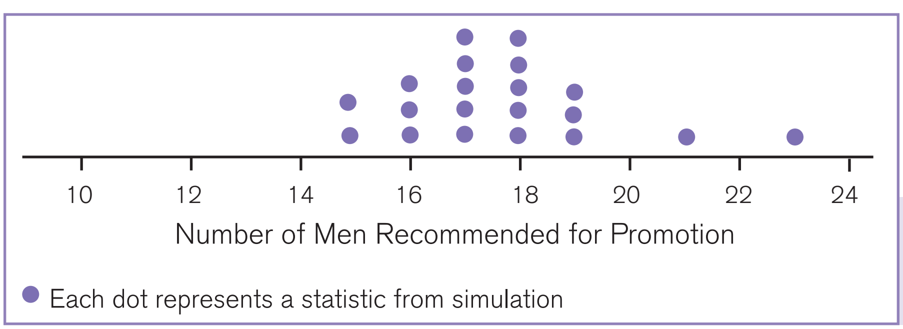


Fig. **2.2.**

The number of men recommended for promotion in 20 more simulations

Statisticians would formalize the question of interest in Discrimination or Not? in two statements: *a null hypothesis,* which would assume that discrimination played no part in the recommendations and that any departure from table 2.1 was due solely to chance variation, and an *alternative hypothesis,* which would assume that discrimination against women played a role in the outcome.

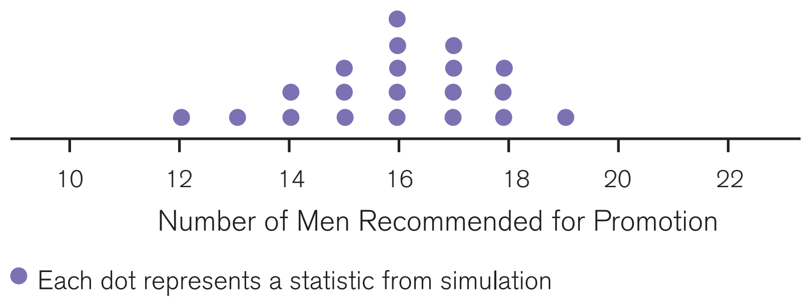
pp. 104, 109–12

“All students should … select and use appropriate statistical methods to analyze data for univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics.”

(NCTM 2000, p. 324)

Fig. **2.3.**

The number of men recommended for promotion in another 20 simulations



Recall that the *interquartile range* is the distance covered by the middle 50 percent of the values of a data set. To calculate the interquartile range (IQR), first order the data from lowest to highest values. Then divide the set into two equal groups at the "middle." If the set contains an even number of values, divide it at the median. If the set contains an odd number of values, make your groups from the points on either side of the median. Next, find the medians of the low and high "halves." These medians are the first and third quartiles (25th and 75th percentiles), often denoted as Q_1 and Q_3, respectively.

$$IQR = Q_3 - Q_1$$

The sample *standard deviation* (SD) of a data set is determined by–

- finding the difference between each observation and the sample mean,
- squaring the differences to eliminate negative values,
- summing the squares of the differences,
- dividing the total by the number of values in the set *minus one,* and
- taking the square root of the quotient.

We can express this calculation as

$$\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n-1}},$$

where the set has n data values, x_1, x_2, …, x_n, and $\bar{x}$ = the sample mean.

18 males and in fact can be calculated as 17.6 males. The counts vary from 15 to 22; however, most of the observations fall between the values of 16 and 19. The variability in the observations can be described by the interquartile range (IQR), or 2; the sample standard deviation (SD), or 1.6, or the sample mean absolute deviation (MAD), or 1.2 males.

The simulated sampling distribution in figure 2.1 conforms to what we might expect to occur in the selection process if there were no discrimination. If the process of selecting applicants for promotion occurred as modeled by the cards, then we would expect about half the promotions to be male and the other half to be female. Therefore, out of 35 cards, the expected number of black cards would be 17.5. The mean of the 20 sample counts was 17.6. The simulated sample distribution is approximately symmetrical to a vertical line through the center of the distribution.

The results of the simulations in figure 2.1 do seem to provide evidence that the selection of 21 males for promotion was not due to chance variation with no discrimination. Only 1 result of the 20 simulated counts was 21 or higher. Thus, for these data, an estimate of the probability of obtaining 21 or more black cards would be 1/20, or 5 percent. The selection process modeled by the cards was chance variation with no discrimination. Is 5 percent small enough to support a claim of discrimination? Do the simulated results in figure 2.1 supply enough evidence to support the possibility that selecting 21 males out of 35 recommended for promotion was not due merely to chance but instead was due to discrimination against women?

Figure 2.2 shows a second set of 20 simulated counts. On the basis of these 20 simulated results, the estimated chance of obtaining 21 or more black cards would be 2/20, or 10 percent. The card selection process represents chance variation with no discrimination. Is 10 percent small enough to support a claim of discrimination against females? How small is "small enough"? What intuitively is "small enough"? Most people are comfortable about saying that an event that occurs no more often than 1 out of 20 times, or 5 percent of the time, is a rare event. Most people would be less confident about calling an event that occurs as often as 2 out of 20 times, or 10 percent of the time, rare.

How small is "small enough"?

You can use two simple exercises to help give your students a sense of "how small is small enough"—that is, how small the probability of an event occurring needs to be for the event to be considered rare. The first exercise calls for two new decks of cards in sealed boxes. Before class, you should—

- open the boxes carefully (so that they can be resealed) and remove the cards, separating red and black cards;
- put all the red cards (including jokers) in one box and all the black cards (including jokers) in the other box;
- reseal the boxes, noting which box has the black cards and which one has the red cards;
- take the box containing the red cards to class and offer a prize for anyone who draws a black card;
- break the seal in front of the class and remove the jokers (reinforcing the illusion of a sealed, intact, new deck of cards);
- direct the students to line up to draw a card. (To suggest that you are keeping the probability of drawing a black card constant for each draw, replace the card that each student draws and reshuffle the deck.).

It usually takes only about five students drawing red cards one after the other for a class to suggest that the deck does not contain any black cards. Since a student is drawing a card from a 52-card deck, the probability of drawing a black card should be 26/52, or 1/2, for each draw. The probability of drawing 5 red cards in a row is $(1/2)^5$, or 3.125 percent. Students will suspect that observing 5 red cards in a row is not likely to happen by chance alone; the probability that you would observe 5 red cards in a row is less than 5 percent.

The second exercise uses a two-headed coin to help students develop a sense of how infrequent an event has to be to be considered rare. Two-headed coins are available in kits of magic tricks or in the shops that sell them. For this exercise—

- bring such a coin to class and hold it up for the students to see, being careful not to reveal that both sides are heads;
- call heads, and then toss the coin, asking for a student to call tails;
- flip the coin in the same manner several times, reporting the results each time;
- Keep flipping the coin until the students conclude that heads appear on both sides.

Usually, as with the cards, this process takes only about five turns. If a coin is fair (head on one side, tail on the other, with each side having a 50 percent chance of landing face up after a toss), the probability of observing 5 heads in a row would be $(1/2)^5$ or 3.125 percent. Once again, most students will intuitively suspect that something rare is happening when five tosses produce 5 heads in a row. The probability of this happening by chance is again less than 5 percent, which in fact is a typical benchmark that statisticians use for calling an event rare. That is, in statistics, an event is often considered rare if it happens by chance 5 percent of the time or less. (In some situations, 5 percent may not be appropriate; the situation may help dictate the percentage.)

Figure 2.3 displays a third set of the counts of the number of men recommended for promotion from 20 simulations. The simulated sampling distribution here is approximately symmetrical, with a slight tail (*skewness*) to the left. Most of the counts are from 15 to 18. There are no counts higher than 19. As estimated in this chance process, if no

The sample *mean absolute deviation* is the average distance of the values in a data set from the mean. Expressed mathematically, the mean absolute deviation of a data set is equal to

$$\frac{|x_1-\bar{x}|+|x_2-\bar{x}|+\ldots+|x_n-\bar{x}|}{n},$$

where the set has n values, x_1, x_2, … x_n, and x is the sample mean.

The probability of observing 21 or more black cards if the selection process is due to chance variation is called the *p*-value. Small *p*-values provide strong support for the alternative hypothesis—in our situation, the hypothesis that discrimination against women could have played a role in the selection process.

"All students should … develop and evaluate inferences and predictions that are based on data by using simulations to explore the variability of sample statistics from a known population." *(NCTM 2000, p. 324)*

The probability that is set as the boundary for calling an occurrence a rare event is called the *significance level.*

discrimination were occurring, the probability of observing 21 or more males would be 0. This simulated sampling distribution provides strong evidence that the selection of 21 or more males recommended for promotion was not due to chance. Note that the mean and median of the simulated sampling distribution are approximately 16, lower than expected and different from the centers in figures 2.1 and 2.2. These results could have occurred as a result of the random nature of dealing the cards, but a possibility also exists that the student carried out the simulation improperly.

Instances of 21, 22, or 23 females could happen by chance variation with no discrimination against women. Having a disproportionate number of females is not out of the ordinary for this chance process. What is unusual here is where the distribution is centered.

How stable are the results?

Figures 2.1, 2.2, and 2.3 illustrate the fact that the simulated sampling distributions of 20 simulated counts of the number of men recommended for promotion will fluctuate from one simulated sampling distribution to another. Combining all the student simulations into one simulated sampling distribution will stabilize the sampling distribution of counts and help students understand what can be expected to happen in the long run. Figure 2.4 illustrates a data set of 1291 student simulations.

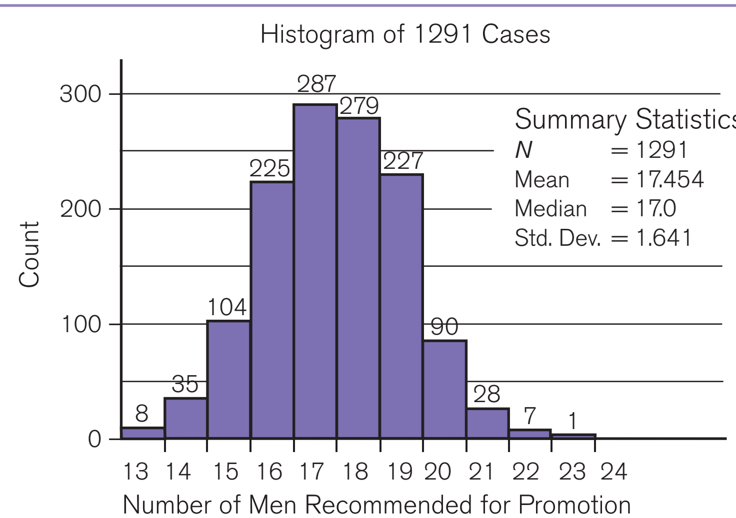
"All students should … understand and apply basic concepts of probability by using simulations to construct empirical probability distributions." (NCTM 2000, p. 324)

This larger set of simulated counts provides a relatively stable model of the shape of the sampling distribution of the number of black cards selected out of 35 to represent the number of men recommended for promotion. The shape is approximately symmetrical, centering at a median of 17 and a mean of approximately 17.5. The variability is similar to that observed in the previous simulated sampling distributions, with an interquartile range of 3 and a standard deviation of about 1.6.

Equipped with the information from the 1291 simulations, you and your students can now reconsider the original question: Is there evidence of possible discrimination against females? Does there appear to be evidence to support a claim that the recommendation of 21 males out of 35 candidates for promotion (black cards) was not due to chance variation? Figure 2.4 displays 36 out of 1291 counts that were 21 or

Fig. **2.4.**

The number of men recommended for promotion in 1291 simulations



higher. Thus, if no discrimination were occurring, the simulated probability of observing 21 or more black cards (that is, the chance of recommending 21 or more males for promotion) would be 36/1291, or 2.79 percent, assuming that dealing the 35 cards was a random process.

This result would seem to provide evidence to support a claim that recommending 21 men out of 35 candidates was not due to chance variation but could have been due instead to discrimination against women. However, it is worth noting that if the number of men recommended for promotion in the study had been 20—just one fewer than the actual number of 21—we would be trying to answer the question of interest by looking at the simulated probability of observing 20 or more black cards, or 126/1291, which is 9.52 percent. This percentage is obviously substantially higher than the selected benchmark of 5 percent for a rare event, and in such a case, we would be less sure about advancing a claim of discrimination against women.

How reliable are the results?

How reliable or consistent are the simulated class results from one class of students to another? How do they compare to previous results? Breaking down the class data set in figure 2.4 into the two school terms from which the results were obtained illustrates the reliability of the simulated results. Figure 2.5 displays the simulated sampling distributions of the results for the two different groups of students who participated in this activity.

Even though figure 2.5a has fewer simulated results than figure 2.5b (278 compared to 1013), the shapes of the two simulated sampling distributions are very similar, and both are approximately symmetrical. They are both centered at a mean of approximately 17.5. Measures of the variability of the two simulated sampling distributions are also similar, with interquartile ranges of 3 (from 16 to 19) for both and standard deviations of 1.62 and 1.64. In figure 2.5a, the chances of obtaining 21 or more black cards were 8 out of 278, or 2.9 percent. In figure 2.5b, the chances were 28 out of 1013, or 2.76 percent. The simulated results appear to be reliable and consistent—that is, the results are repeatable.

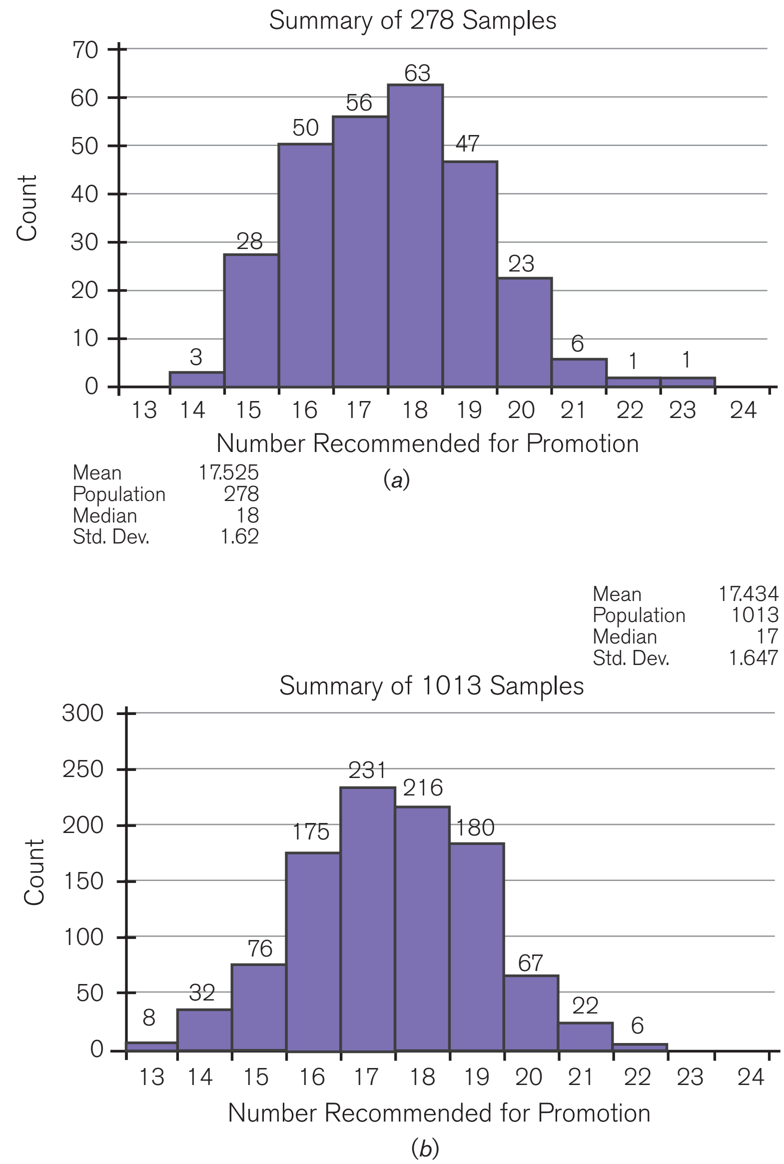
Conclusion

This chapter's examination of the scenario Discrimination or Not? focused on making a decision about discrimination in a situation where there was variability and uncertainty. The question was whether the outcome observed in the study—21 men recommended for promotion out of 35 recommended candidates—could reasonably be attributed to chance variation or was more likely to be due to some other effect, such as discrimination against female candidates. By exploring the data graphically and using numerical summaries and simulated probabilities, you and your students will have discovered that you can deal statistically with the variability (uncertainty) of the data and make a statistical decision about the case.

Two possible errors can enter into the decision-making process of statistical inferences. Remember that the reasoning behind any claim of discrimination against females in the bank supervisors' recommendations rests on the fact that the likelihood of actually recommending 21 or

Fig. **2.5.**

The number of men recommended for promotion in (a) 278 simulations and (b) 1013 simulations



See the CD-ROM for a discussion of probability theory related to the scenario.

A discussion of both types of errors—type I and type II—is included on the CD-ROM.

more males if no discrimination were occurring is small (less than 5 percent). What if, however, in the case of the study, the recommendations of the 21 male candidates happened to be part of the approximately 5 percent of such results that occur by chance? On the one hand, we could potentially make a mistake in claiming that discrimination against women played a role in the bank supervisors' decisions when it did not. This would be an example of a *type I error*. Errors of this type occur when the null hypothesis—in our case, that there was no discrimination—is rejected but actually is true.

On the other hand, we could potentially make a mistake of a different type—*a type II error*—by not supporting a claim of discrimination against women when such discrimination existed. This type of error occurs when the null hypothesis—in our situation, the hypothesis that there was no discrimination—*should* be rejected in favor of the alternative hypothesis—here, that there was discrimination against women—but is not.

At the beginning of chapter 2, the activities introducing the scenario Discrimination or Not? raised questions about the design of the study. What important assumptions were made, and how did they affect the outcome? What elements were necessary to make this a well-designed study? One assumption was that the recommendation of each bank supervisor was an individual decision and that supervisors did not discuss the files with one another or think that other supervisors would be looking at the files. For statistical purposes, there was a need to assume *independence* among the bank supervisors (which actually existed in the original study).

Other questions are also relevant. How were the files assigned to particular supervisors? Would it have been important to use a random process in deciding whether a supervisor got a file labeled "male" or a file labeled "female"? Why, or why not? How might statistical decisions about possible discrimination have been affected if more than 48 supervisors had participated in the study?

In the actual study, except for the genders assigned to the candidates, all the files contained the same information. In a real-life situation, it is likely that there would be 48 completely different files under consideration. It would not be reasonable to assume that the files represented equally qualified candidates. A statistical analysis would have to find a way to incorporate the differences among the applicants. If an actual discrimination suit against a bank were being tried in a court of law, a jury would need to consider, among other things, whether the bank had a history of discriminatory practices.

The questions and statistical issues raised in this chapter will be explored further in the following chapters. It is important to keep in mind that in all the activities involving the scenario Discrimination or Not? statistical analysis does not prove the presence or absence of discrimination against women; it only sheds light on expected behavior under chance variation when no discrimination has occurred.