Introduction: Most statistics textbooks begin with a detailed look at either data analysis or data production. While there are compelling arguments for beginning with either topic, both approaches leave students without much insight about what statistics is all about. An intriguing alternative is to start by examining a statistical study from the original research question to the data production, data analysis, probability models, and inferential thinking that leads to a conclusion.

This investigation is designed to be used early in the course to give students an overview of the statistical problem solving process. Researchers start with a question of interest: are drivers more distracted when using a cell phone than when talking to a passenger in the car? To help answer their question, researchers design a plan for producing data, and then carry out the plan. With data in hand, students are asked to assist with preliminary graphical and numerical data analysis. Students will observe that a much higher proportion of drivers who are talking on a cell phone miss a designated freeway exit than drivers who are talking to a passenger. Is this difference a direct result of the difference in experimental conditions—talking on a cell phone versus talking to a passenger? Or could the difference be explained by the chance involved in assigning people to the two groups when there is actually *no difference* in the effects of the two experimental conditions on drivers' distraction? To find out, students will use a physical simulation with a deck of cards and the results of a computer simulation to estimate the likelihood that a difference in proportions equal to or greater than the one in the actual experiment would occur "just by chance." Based on the resulting probability estimate, students will draw a conclusion about which is the more likely explanation for the difference. In a 45-50 minute period, students can get a glimpse of how the four primary themes of AP* Statistics—data production, data analysis, probability models, and inference—fit together in answering an engaging research question.

Content notes: Many statistics textbooks fail to make a clear distinction between inference based on sampling situations and inference based on experiments. In a sampling setting, random sampling allows us to generalize sample results to a larger population of interest. The individuals in an experiment are usually *not* chosen at random from some larger population. In a well-designed comparative experiment, individuals are randomly assigned to experimental conditions. This random assignment helps create roughly equivalent groups before the experimental conditions (treatments) are administered. If substantial differences in the response variable emerge between the different treatment groups, we can attribute that difference in responses to the difference in treatments. (After all, if the groups were roughly equivalent to begin with, and the only difference is the treatment administered to the two groups, then any difference in responses should be due to the difference in treatments.) In that case, we say that there is evidence of a "treatment effect." Put another way, random assignment in experiments helps us draw cause-and-effect conclusions.

The other important role of random sampling and random assignment is to introduce chance into the process. Sampling distributions are the probability models that describe how sample statistics vary in repeated random samples from the same population. Randomization distributions are the probability models that describe how the difference in sample statistics varies in repeated random assignments of subjects to the two treatment groups in an experiment when there is no difference in the effects of the treatments on subjects' responses.

Set-up: Divide students into groups of 3 or 4 for this investigation. Give each group a standard deck of playing cards and enough copies of the Distracted Driving student investigation for each member of the group.

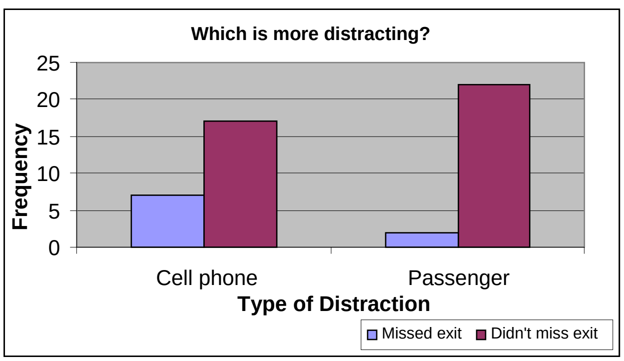
Are drivers more distracted when using a cell phone than when talking to a passenger in the car? Researchers wanted to find out, so they designed an experiment. Here are the details.

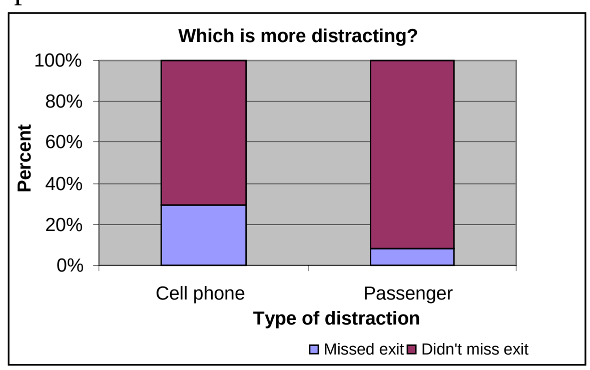
In a study involving 48 people, 24 people were randomly assigned to drive in a driving simulator while using a cell phone. The remaining 24 were assigned to drive in the driving simulator while talking to a passenger in the simulator. Part of the driving simulation for both groups involved asking drivers to exit the freeway at a particular exit. In the study, 7 of the 24 cell phone users missed the exit, while 2 of the 24 talking to a passenger missed the exit. (from the 2007 AP* Statistics exam, question 5)

- Let's start by summarizing the data from this study. Each of the 48 people in the experiment can be classified into one of the four cells in the table below based on the experimental condition to which they were assigned and whether they missed the designated exit. Use information from the previous paragraph to complete the table.

		Distraction	
		Cell phone	Passenger
Missed exit?	Yes	7	2
	No	17	22

To analyze data, we begin by making one or more graphs.





- Two types of Excel bar graphs are shown above. Explain the difference in what the two graphs display. Then tell which one you prefer and why.

  The left-hand comparative bar graph compares the *counts* of drivers who missed the exit in the two groups (light bars) and the *counts* of drivers who didn't miss the exit in the two groups (dark bars). The right-hand segmented bar graph compares the *percents* who missed the exit in the two groups (light shading = missed exit; dark shading = didn't miss exit).

  Since the two groups in this experiment are the same size, it makes sense to compare either counts or proportions. Which graph you like better is a matter of personal preference.

Next, we add numerical summaries. We might be interested in comparing the counts, percents, or proportions of people in the two groups who missed the freeway exit.

- Fill in the missing entries in the table below for the passenger group.

	Missed exit		
Group	Number	Proportion	Percent
Cell phone	7	0.292	29.2
Passenger	2	0.083	8.3

Caution! These next two paragraphs describe the logic that underlies making an inference about cause and effect in this experimental setting. Take time to review the key ideas with your students before you have them complete the hands-on activity that follows.

- *Either the two experimental conditions—talking on a cell phone and talking to a passenger—had a differential effect on whether drivers missed the exit or they had the same effect on whether drivers missed the exit. Researchers believe that the two experimental conditions will show a differential effect on drivers' distraction, but the burden of proof is on them to provide compelling evidence of this.*
- *Is it plausible that the two experimental conditions had the same effect on drivers' distraction? To find out, we assume that each driver would have the same result in terms of missing the freeway exit or not whether they were in the cell phone group or the passenger group. Then we reassign the subjects to the two groups at random and count how many people miss the exit in each group. If we do this "random reassignment" many times, we can see how often the difference in proportions for the two groups equals or exceeds the difference observed in the actual experiment "just by chance" when there is no differential treatment effect. Suppose we rarely get a difference as large as 0.209 in many random reassignments. In that case, it doesn't seem plausible that the two experimental conditions could have had the same effect on drivers' distraction. However, if our random reassignments frequently yield results as extreme as or more extreme than those in the actual experiment, it is plausible that there is no difference in treatment effects.*

In the distracted driving experiment, 29.2% of the 24 drivers talking on cell phones missed the freeway exit, compared with only 8.3% of the 24 drivers who were talking to passengers. This seems like a pretty large difference—almost 21% higher for the drivers who used cell phones. Researchers might be tempted to conclude that the different experimental conditions—talking on a cell phone and talking to a passenger—actually caused the observed difference in the percent of drivers who missed the freeway exit. There is another possibility, however.

Suppose that the two experimental conditions—talking on a cell phone and talking to a passenger—actually have *the same effect* on drivers' distraction. In that case, the 9 people in this experiment who missed the freeway exit would have done so no matter which group they were assigned to. Likewise, the 39 people who did not miss the exit would have had the same result whether they talked on a cell phone or to a passenger. This leads us to the other possibility: if the two experimental conditions actually have the same effect on drivers' distraction, then the difference in the percents that missed the exit in the two groups could simply have been due to chance. That is, the difference could be a result of which 24 people just happened to be assigned to each group. In the next activity, you will examine whether this second possibility seems plausible.

Activity: Could the observed difference be due to the chance assignment of people to groups?
Materials: Standard deck of playing cards for each group of 3-4 students

What would happen if we reassigned the 48 people in this experiment to the cell phone and passenger groups many times, assuming that the group assignment had no effect on whether each driver missed the exit? Let's try it and see.

1. Get a standard deck of playing cards from your teacher. Make sure that your deck has 52 cards, not including jokers.

2. We need 48 cards to represent the 48 drivers in this study. In the original experiment, 9 people missed the exit and 39 people didn't miss the exit. If the group assignment had no effect on drivers' distraction, these results wouldn't change if we reassigned 24 people to each group at random. For a physical simulation of these reassignments, we need 9 cards to represent the people who will miss the exit and 39 cards to represent the people who won't miss the exit. With your group members, discuss which cards should represent which outcomes. When you have settled on a plan, designate one member of your group to share your plan with the class.

Plans will vary.

3. After each group presents its plan, the class as a whole will decide which plan to use. Record the details here.

Answers will vary. Here's one possibility: Remove the 10, J, Q, and K of clubs from the deck. Let the Ace through 9 of clubs represent missing the exit; all other cards represent not missing the exit.

4. Now you're ready to simulate the process of reassigning people to groups. "Shuffle up and deal" two piles of 24 cards—the first pile representing the cell phone group and the second pile representing the passenger group. Record the number of drivers who missed the exit in each group.

Answers will vary for this question and the next.

5. Repeat this process 9 more times so that you have a total of 10 trials. Record your results in the table provided.

Trial	Number who missed exit in cell phone group	Number who missed exit in passenger group
1		
2		
3		
4		
5		
6		
7		
8		
9		

10		

In the original experiment, 7 of the 24 drivers using cell phones missed the freeway exit, compared to only 2 of the 24 drivers who were talking to a passenger. How surprising would it be to get a difference this large or larger simply due to chance if the effects of the two experimental conditions on drivers' distraction were actually the same? You can estimate the chance of this happening with the results of your simulation.

6. In how many of your 10 simulation trials did 7 or more drivers in the cell phone group miss the exit? Why don't you need to consider the number of people in the "talking to a passenger group" who missed the exit?

The two group sizes are fixed at 24 each. The number of drivers who miss the exit is also fixed at 9, because we are assuming that which experimental condition each driver receives will not affect whether that person misses the freeway exit. Once we know the number of people who miss the exit in the cell phone group, that determines the number of people who miss the exit in the passenger group and the resulting difference in the proportions who miss the exit.

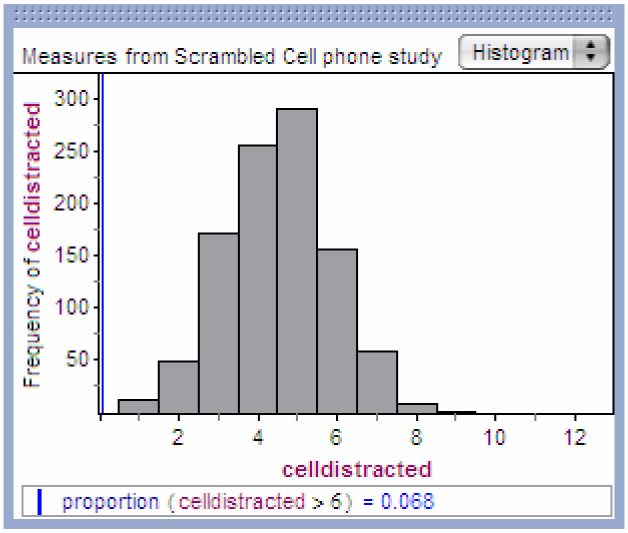
7. Combine results with your classmates. In what percent of the class's simulation trials did 7 or more people in the cell phone group miss the freeway exit?

Answers will vary.

8. Based on the class's simulation results, do you think it's possible that cell phones and passengers are equally distracting to drivers, and that the difference observed in the original experiment could have been due to the chance assignment of people to the two groups? Why or why not?

The theoretical probability of 7 or more drivers missing the exit in the cell phone group when there is no differential treatment effect is 0.068. If the class's simulation results come reasonably close to the theoretical probability, it is certainly plausible that the difference in proportions observed in the experiment could be due to the random assignment of subjects to the two groups. In 100 random reassignments, we would expect a result as extreme as or more extreme than the one in the actual experiment about 7 times just by chance. Students could certainly argue that something which has only a 7% chance of happening isn't plausible enough for their liking. This "threshold" of how unlikely something needs to be before it is implausible provides motivation for the later study of significance levels in hypothesis testing.

Here are the results of 1000 trials of a computer simulation, like the one you did with the playing cards, showing the number of drivers who missed the exit in the cell phone group.



9. In the computer simulation, how often did 7 or more drivers in the cell phone group miss the exit when there is no difference in the effects of the experimental conditions? Do you think the results of the original experiment could be due to chance, and not to a difference in the effects of cell phone use and talking to a passenger on driver distraction? Explain your reasoning.

From the bottom line of the figure, in about 6.8% of random reassignments of the 48 people to the two groups, 7 or more drivers in the cell phone group missed the exit.

If you have questions or suggestions related to this activity, please contact one of the authors:

Daren Starnes	*The Lawrenceville School*	*dstarnes@lawrenceville.org*
Roxy Peck	*Cal Poly San Luis Obispo*	*rpeck@calpoly.edu*
Celia Rowland	*Enloe High School*	*crowland@wcpss.net*