The objective of many statistical studies is to use sample data to tell us something about the population from which the sample was selected. As we have seen in the previous sections, if the sample is selected in an appropriate way and involves some type of random selection, it may be reasonable to regard the sample as representative of the population of interest. Suppose that 15 in a sample of 50 randomly selected students favor banning soda machines from campus—a proportion of 15/50 = .30 or, equivalently, 30%. Is it also reasonable to say that exactly 30% of the students at your school favor banning soda machines from campus? The value of the sample proportion depends on the outcome of the random selection—which 50 students are selected to participate. As a consequence, the value of the sample proportion will vary from one random sample to another, and we can't expect that the sample proportion will be exactly equal to the actual value of the population proportion.

Does this mean the sample proportion doesn't tell us anything about the population proportion? Fortunately, the answer to that question is no! If the sample has been chosen appropriately and can be regarded as representative of the population, we can expect that the sample value will, on average, be "close" to the true population value. But, to really tell us something useful about the population, we need to be a bit more specific about what we mean by "close."

In this section, we will begin to investigate how we can use data from a well-designed study to draw conclusions about a population. This requires an understanding of the nature of **sampling variability**—the chance differences that occur from one random sample to another as a consequence of random selection.

Statistical inference is the process of drawing conclusions about a population using data from a sample. In most statistical studies, sample data is collected so that the investigator can either estimate some population characteristic of interest (such as the proportion of students at your school who favor the ban of soda machines on campus) or evaluate the plausibility of some claim that has been made about the population (for example, a claim that more than 50% of the students at the school favor a ban on soda machines).

In an **estimation** situation, we need to understand sampling variability to be able to assess how close an estimate from a sample is likely to be to the actual value of the corresponding population characteristic. In published reports, you will often see statements that include a **margin of error**. For example, the analysis of data from a survey on public support for the president might include a statement such as "The proportion of U.S. adults who believe the president is doing a good job is .45 (45%) with a margin of error of .03 (3%)." The reported margin of error acknowledges that the true population proportion is not likely to be *exactly* .45 and indicates that, based on what was seen in the sample, plausible values for the population proportion might be anything between .42 and .48. How is the margin of error computed? It is based on an assessment of sampling variability. Investigation 13 illustrates the reasoning that leads to a margin of error calculation.

> What is sampling variability? It refers to the chance differences that occur from one random sample to another.

The second way data from a sample is typically used to draw a conclusion about a population is in the evaluation of the plausibility of a claim about the population. For example, the *Youth Monitor* report titled "Coming of Age in America: Part IV – The MySpace Generation" (*www.greenbergresearch.com*) includes the following statement:

> Half (52%) of 18–24 year olds report they have a page on MySpace. A third report membership on Facebook, though that number rises to over half (54%) among students.

What about students at your school? Does a majority (more than 50%) have a page on MySpace? Suppose the proportion in a random sample of students from your school who report that they have a MySpace page was .52. Can you conclude that a majority of *all* students at the school have a MySpace page? This requires some careful thought. There are two reasons why the sample proportion might have been larger than .50. One reason is sampling variability—we don't expect the sample proportion to be exactly equal to the population proportion. So, maybe the population proportion is .50 (or maybe even something smaller) and a sample proportion of .52 is "explainable" just due to sampling variability. If this is the case, we can't interpret the sample proportion of .52 as convincing evidence that a majority of *all* students at your school have a MySpace page. However, another reason the sample proportion might have been greater than .50 is that the true population proportion is, in fact, greater than .50. Is .52 enough larger than .50 that the difference can't be explained as being due to sampling variability alone? If so, we would say that, based on the sample data, there is convincing evidence that more than half of the students at the school have a MySpace page. How do we make this determination? Again, it is based on an assessment of sampling variability. Investigations 14 and 15 illustrate the reasoning that enables us to use data from a random sample to evaluate the plausibility of a claim about a population.

The investigations in this section are designed to introduce you to reasoning about sampling variability in each of these two types of settings—estimation and evaluating the plausibility of a claim. In this brief introduction, we can illustrate the reasoning, but for a more complete treatment of the inferential process, we encourage you to consider taking a course devoted to statistics and data analysis!

Investigation #13: The Internet —Information or Social Highway?

The report "Coming of Age in America: Part IV—The MySpace Generation" referenced in the Overview to this section also included the following:

> For many young people, the Internet represents not just an information super-highway but, indeed, a social highway. In fact, nearly two thirds (64%) agree that "I don't know how I would keep up with my friends or family if I didn't have the Internet."

What proportion of the students at your school agree with the statement above? In this investigation, you will carry out a simple survey and then use the resulting data to estimate the proportion who agree. You will also see how we can obtain a conservative margin of error for your estimate.

Let's begin by taking a look at margin of error. A **conservative estimate of the margin of error** when a proportion from a random sample is used as an estimate of the corresponding population proportion is $\frac{1}{\sqrt{n}}$, where n denotes the sample size. For a sample size of 40, it is reasonable to use $\frac{1}{\sqrt{n}}$ as a conservative estimate of margin of error for sample proportions between .15 and .85. For larger sample sizes, this way of computing a conservative margin of error is reasonable for an even larger range of values of the sample proportion. When we say that this is a conservative estimate of the margin of error, we mean that the actual margin of error will be equal to or smaller than this conservative value. That is, when we use this conservative estimate, we may be overstating our potential error. This is generally considered to be a better alternative than understating potential error. It is possible to obtain more precise estimates of the margin of error, but that is beyond the scope of this introductory investigation.

To see why it is reasonable to use $\frac{1}{\sqrt{n}}$ as a conservative estimate of the margin of error, we need to consider the role of sampling variability and how the sample proportion varies from one random sample to another. To do this, we will carry out a simulation.

How much will sample proportions vary from one random sample to another and how much will these sample proportions tend to differ from the actual value of the corresponding population proportion? Suppose that the proportion of students at your school who have some characteristic of interest is actually .60. Think of this as a population where 60% have the characteristic we are interested in (called them "successes"). What kind of sample proportions would you expect to see for random samples of size 40 from this population? We can find out by creating a hypothetical population with 60% "successes." One way to create such a population is to use random digits. Because each of the digits 0, 1, 2, …, 9 is equally likely to occur in a list of random digits, if we think of the digits 1, 2, 3, 4, 5, and 6 as representing individuals who are successes and the digits 0, 7, 8, and 9 as representing individuals who are not successes, we can view a collection of random digits as a population with 60% successes.

We can now take a random sample of size 40 from this hypothetical population by selecting 40 random digits. Each member of the sample (each digit) can then be classified according to whether it is a success, and then the sample proportion of success can be computed. For example, consider the list of 40 random digits below.

0 5 2 3 5 7 2 3 2 3 0 7 1 1 0 0 3 2 5 8 1 2 4 6 9 4 5 3 1 7 0 2 1 5 9 3 9 4 7 6

Identifying the 1s, 2s, 3s, 4s, 5s, and 6s as successes (the underlined digits above), we could then compute the proportion of successes for this random sample:

$$\text{sample proportion} = \frac{\text{number of successes in the sample}}{n} = \frac{27}{40} = .675$$

Repeating this process a large number of times and noting the values of the sample proportions obtained allows us to see what values of the sample proportion are expected as a result of sampling variability alone.

1. Use a sequence of random digits to represent a random sample from a population that has 60% successes (a population proportion of .60). Identifying successes as described in the paragraph above, determine the number of successes in the sample and the sample proportion.

Number of successes in the sample =

Proportion of successes for this sample =

2. Record the sample proportion obtained in step 1 in the table below. Then, repeat the process four more times. Enter each of the four new sample proportions into the table below.

Sample Number	1	2	3	4	5
Sample Proportion					

3. Enter a tally mark for each of your observed sample proportions in the table on the following page. Then enter a tally mark for each of the five sample proportions observed by other students in your class. You can do this by making tally marks as each student reads his or her sample proportions. When finished, count the tally marks for each row in the table and enter the total in the frequency column for the appropriate row.

4. Use the information in the table from step 3 to complete the following sentences.

(a) About 2.5% of the observed sample proportions are less than _____.

(b) About 2.5% of the observed sample proportions are greater than _____.

(c) About 95% of the observed sample proportions are between _____ and _____.

Sample Proportion	Tally	Frequency	Sample Proportion	Tally	Frequency
0.000			0.525		
0.025			0.550		
0.050			0.575		
0.075			0.600		
0.100			0.625		
0.125			0.650		
0.150			0.675		
0.175			0.700		
0.200			0.725		
0.225			0.750		
0.250			0.775		
0.275			0.800		
0.300			0.825		
0.325			0.850		
0.350			0.875		
0.375			0.900		
0.400			0.925		
0.425			0.950		
0.450			0.975		
0.475			1.000		
0.500					

5. Now try to complete the following statement:

> About 95% of the time, the sample proportion was within _____ of the actual population proportion (.60).

Hint: Look at your answer to part (c) of question 4. How far are the endpoints of the interval you give there from .60?

6. Compute the conservative margin of error for a sample size of 40.

Conservative margin of error = $\dfrac{1}{\sqrt{n}}$ =

7. How does the conservative margin of error from step 6 compare to your number from step 5?

What if the proportion of successes in the population was .2 instead of .6? What kind of sample proportions should we expect to observe? We can investigate this by carrying out a simulation that is similar to the one just carried out. Only a few modifications are needed.

8. If a sequence of random digits is to be used to represent a random sample from a population that has 20% successes, what digits will you use to represent successes in the sample?

9. Use a sequence of random digits to represent a random sample of size 40 from a population that has 20% successes (a population proportion of .20). Determine the number of successes in the sample and the sample proportion.

Number of successes in the sample =

Proportion of successes for this sample =

10. Record the sample proportion obtained in step 9 in the table below. Then repeat the process four more times. Enter each of the four new sample proportions into the table.

Sample Number	1	2	3	4	5
Sample Proportion					

11. Enter a tally mark for each of your observed sample proportions in the table on the next page. Then enter a tally mark for each of the five sample proportions observed by other students in your class. You can do this by making tally marks as each student reads his or her sample proportions. When finished, count the tally marks for each row in the table and enter the total in the frequency column for the appropriate row.

12. Write a few sentences describing how the observed sample proportions for random samples of size 40 differ for a population with 60% successes and a population with 20% successes.

Sample Proportion	Tally	Frequency	Sample Proportion	Tally	Frequency
0.000			0.525		
0.025			0.550		
0.050			0.575		
0.075			0.600		
0.100			0.625		
0.125			0.650		
0.150			0.675		
0.175			0.700		
0.200			0.725		
0.225			0.750		
0.250			0.775		
0.275			0.800		
0.300			0.825		
0.325			0.850		
0.350			0.875		
0.375			0.900		
0.400			0.925		
0.425			0.950		
0.450			0.975		
0.475			1.000		
0.500					

13. Use the information in the table to complete the following sentences.

(a) About 2.5% of the observed sample proportions are smaller than _____.

(b) About 2.5% of the observed sample proportions are greater than _____.

(c) About 95% of the observed sample proportions are between _____ and _____.

14. Now try to complete the following statement:

About 95% of the time, the sample proportion was within _____ of the actual population proportion (.20).

Hint: Look at your answer to part (c) of question 13. How far are the endpoints of the interval you give there from .20?

15. How does the conservative margin of error for a sample of size 40 (see step 6) compare to your number from step 14?

16. Consider the following statement from the introduction to this investigation:

> When we use this conservative estimate, we may be overstating our potential error. This is generally considered to be a better alternative than understating potential error. It is also worth noting that the conservative estimate overstates to a larger degree the more that the sample proportion differs from .5.

Is this statement supported by the two simulations that you have carried out? Explain.

Now that you understand what is meant by conservative margin of error and know how to compute it, you are ready to see how this is used when reporting survey data. Recall the question posed at the beginning of this investigation:

> What proportion of the students at your school agree with the statement "I don't know how I would keep up with my friends or family if I didn't have the Internet"?

17. As a class, discuss how you will go about selecting a sample of students at your school to participate in a survey. Remember that the sample should be selected in such a way that it will be reasonable to generalize from the sample to the population of all students at the school. Be sure to consider the ethical issues involved in conducting a survey—informed consent, right of refusal, and preserving anonymity and confidentiality. Write a brief summary of your plan for carrying out the survey.

18. Once your teacher has approved your plan, carry out the sampling that is required. Then, ask each individual selected for inclusion in the sample whether he or she agrees with the statement "I don't know how I would keep up with my friends or family if I didn't have the Internet." Use the resulting data to compute the sample proportion and a conservative margin of error.

Sample size =

Number in sample who agree with statement =

Sample proportion =

Conservative margin of error =

19. Write a paragraph describing the results of your survey. Be sure to reference the estimate of the proportion of students at your school who agree with the statement and the margin of error. Also, indicate whether you think the statement from the report referenced at the beginning of this investigation is accurate for your school and why or why not. (The report said nearly two-thirds of young people agree with the statement.)

Investigation #14: Evaluating the MySpace Claim

In the Section IV Overview, the following question was posed: Does a majority (more than 50%) of students at your school have a page on MySpace? In this investigation, you will carry out a survey of 60 students at your school. Then, you will use the resulting data to determine if there is convincing evidence that a majority of students at your school have a MySpace page.

Before we collect data, let's consider what it would take to convince us that a majority of students at the school have a MySpace page. Suppose we select a random sample and find that the proportion in the sample who report they have a MySpace page is less than .50 (less than 50%). That certainly wouldn't be convincing evidence that more than 50% have such a page. But what if our sample proportion is greater than .50? Should we be "convinced" that a majority of *all* students (not just those in the sample) have a MySpace page? This decision isn't so easy!

What we need to see to be *convinced* is not just a sample proportion greater than .50—it has to be *enough* greater that it is not likely to have occurred just by chance due to sampling variability. That is, to be convinced, we would need to rule out sampling variability—those chance differences that occur from one sample to another as a consequence of random selection—as a plausible explanation for what we see in the sample. For our MySpace example, if we decide that the sample proportion is enough larger than .50 that we don't think it is just due to sampling variability, we would conclude that there is convincing evidence that the actual proportion in the population is larger than .50.
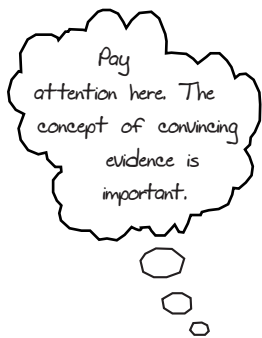
So, how do we decide just how much greater than .50 our sample proportion must be in order to convince us that what we are seeing is not just sampling variability? To figure this out, we need to know what kind of sample proportions are likely to be observed just by chance when the population proportion is .50 or less (not a majority). If we see one of these proportions, we would not want to say we were convinced that a majority of students have a MySpace page.

The first steps in this investigation use simulation to explore what values of the sample proportion would *not* be convincing evidence of a majority. Suppose that in fact 50% or fewer of the students at your school have a MySpace page. We will focus on the most extreme of these possible population values—50%. We use 50% (a proportion of .50) in the simulation that follows because to be convinced that a majority of students have a MySpace page, we need to be convinced that the population proportion is greater than .50.

To determine what sample proportions are likely to be observed when the population proportion is .50, we will use a hypothetical population of random digits. We create a population with 50% "successes" by designating the digits 1, 2, 3, 4, and 5 to represent individuals who are successes and the digits 6, 7, 8, 9, and 0 to represent individuals who are not successes.

Pay attention here. The concept of convincing evidence is important.

We can now take a random sample of size 60 from this hypothetical population by selecting 60 random digits. Each member of the sample (each digit) can then be classified according to whether it is a success. Then, the sample proportion of successes can be computed. For example, consider the list of 60 random digits below.

8 4 6 6 4 3 7 8 6 7 3 0 1 3 5 4 6 7 6 2 6 3 8 2 9 9 1 9 0 0
3 0 6 3 7 7 0 5 9 1 8 1 3 4 3 6 5 4 0 3 0 9 2 2 5 9 7 1 5 9

Identifying the 1s, 2s, 3s, 4s, and 5s as successes (the underlined digits above), we could then compute the proportion of successes for this random sample:

$$\text{sample proportion} = \frac{\text{number of successes in the sample}}{n} = \frac{28}{60} = .467$$

Repeating this process a large number of times and noting the values of the sample proportions obtained allows us to see what values of the sample proportion are expected as a result of sampling variability alone.

1. Use a sequence of 60 random digits to represent a random sample of size 60 from a population that has 50% successes (a population proportion of .50). Identifying successes as described in the paragraph above, determine the number of successes in the sample and the sample proportion.

Number of successes in the sample =

Proportion of successes for this sample =

2. Record the sample proportion obtained in step 1 in the table below. Then, repeat the process four more times. Enter each of the four new sample proportions into the table.

Sample Number	1	2	3	4	5
Sample Proportion					

3. Enter a tally mark for each of your observed sample proportions in the table on the next page. Then, enter a tally mark for each of the five sample proportions observed by other students in your class. You can do this by making tally marks as each student reads his or her sample proportions. When finished, count the tally marks for each row in the table and enter the total in the frequency column for the appropriate row.

4. Use the information in the table from step 3 to complete the following sentence.

About 5% of the observed sample proportions are greater than _____.

5. Based on the statement in step 4, it is reasonable to say that when the true population proportion is .50, a value as large as or larger than the number in step 4 would

Sample Proportion	Tally	Frequency	Sample Proportion	Tally	Frequency
0.017			0.517		
0.033			0.533		
0.050			0.550		
0.067			0.567		
0.083			0.583		
0.100			0.600		
0.117			0.617		
0.133			0.633		
0.150			0.650		
0.167			0.667		
0.183			0.683		
0.200			0.700		
0.217			0.717		
0.233			0.733		
0.250			0.750		
0.267			0.767		
0.283			0.783		
0.300			0.800		
0.317			0.817		
0.333			0.833		
0.350			0.850		
0.367			0.867		
0.383			0.883		
0.400			0.900		
0.417			0.917		
0.433			0.933		
0.450			0.950		
0.467			0.967		
0.483			0.983		
0.500			1.000		

occur only about 5% of the time just by chance. That means that it would be fairly unusual to see a value this large if the population proportion is .50 (and even less likely if the population proportion were something smaller than .50). Use the number from step 4 to complete the following decision statement:

If we take a random sample of 60 students from our school and we find that the proportion who have a MySpace page is equal to or greater than _____, we will conclude that there is convincing evidence that a majority of students at the school have a MySpace page.

6. Write a few sentences explaining why the decision statement in step 5 makes sense.

7. We are almost ready to collect some data. As a class, discuss how you will go about conducting a survey involving 60 students at your school. Remember that the sample should be selected in such a way that it will be reasonable to generalize from the sample to the population of all students at the school. Be sure to consider the ethical issues involved in conducting a survey—informed consent, right of refusal, and preserving anonymity and confidentiality. Write a brief summary of your plan for carrying out the survey.

8. Implement the sampling plan described in the previous step, asking each chosen individual whether he or she has a MySpace page. Use the resulting data to compute the sample proportion.

Sample size =

Number in sample who agree with statement =

Sample proportion =

9. Based on the observed sample proportion and the decision statement from step 5, which of the following possible conclusions is appropriate?

Conclusion 1: There is convincing evidence that a majority of students at our school have a MySpace page.

Conclusion 2: There is not convincing evidence that a majority of students at our school have a MySpace page.

10. Write a paragraph describing your conclusions. Include an assessment of whether it is plausible that 50% or fewer of the students at your school have a MySpace page and that what was observed in the sample can be explained by sampling variability alone.

Investigation #15: Are Teens the Same Everywhere?
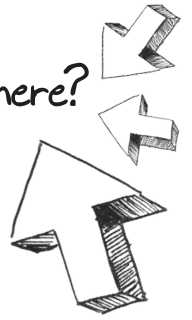
Students from a high school in Milwaukee, Wisconsin, and from a high school in Sweden participated in a survey that included the following seven questions.

(1)   Indicate your gender: _____ Male _____ Female

(2)   You have lots of homework to do tonight. What choice best describes what you will probably do?

   (A)   You do what you want and forget to complete the homework.

   (B)   Start it, but quit when it gets hard.

   (C)   Wait until the last minute, but do it.

   (D)   Do it right away.

(3)   You are scheduled to work, but there is a party you would like to go to. What choice best describes what you will probably do?

   (A)   You go to the party and don't go to work.

   (B)   You call in sick and go to the party.

   (C)   You make an excuse to leave work an hour or two early so you can go to the party.

   (D)   You work your entire shift and only go to the party if it is still on after work.

(4)   Your parents ask you to do the dishes. What choice best describes what you will probably do?

   (A)   You ignore them and don't do the dishes.

   (B)   You say you'll get to them, but never "find the time."

   (C)   You procrastinate, but finally get them done.

   (D)   You do them quickly and without your parents nagging you.

(5)   You have a curfew of midnight. What choice best describes what you will probably do?

      (A)   You don't call and come home when it suits you.

      (B)   You finally call after you are already late and make excuses.

      (C)   You call to tell your parents you will be late.

      (D)   You get home on time.

(6)   The cashier mistakenly gives you an extra $10 in change. What choice best describes what you will probably do?

      (A)   You keep the extra money and spend it.

      (B)   You keep it, but feel guilty about it.

      (C)   You start to keep it, but then decide to return it.

      (D)   You give it back as soon as you realize you have it.

(7)   You have a major paper due in one month. What choice best describes what you will probably do?

      (A)   You never do it.

      (B)   You finish it under pressure after the due date.

      (C)   You stay up late to finish it the night before it is due.

      (D)   You plan ahead and get it done on time without rushing.

A total of 60 students at the Milwaukee school responded to the survey and a total of 56 students from the school in Sweden responded. The sample at each school was selected in a way that was designed to produce a representative sample of the students at that school. The survey data are given in the tables that appear at the end of this investigation.

In this investigation, we will see how the data from these two samples can be used to decide if there is a "significant difference" in the proportion of students at the Milwaukee school and the proportion of students at the school in Sweden who respond in a particular way

to one of these questions. By "significant difference" we mean that the difference in the response proportions for the two samples is larger than we would have expected to see just as a result of sampling variability.

The investigation will first guide you through the steps necessary to answer the following question:

> Is there a significant difference in the proportion of students who report that they would probably get home on time when given a curfew of midnight for students at the two schools?

After seeing how this question is answered, you will then be asked to answer an additional question using a similar process.

1. Let's consider the question posed: Is there a significant difference in the proportion of students who report that they would probably get home on time when given a curfew of midnight for students at the two schools?

To answer this question, we will need to look at the responses to question number 5 on the survey. In particular, we are interested in the proportions of students who answer D to this question. Answers A, B, and C can be combined into a single category here because they all represent answers where the student says he or she will probably not get home on time. For each of the two samples, we then need to know how many students answered D to question 5 and how many gave an answer other than D. We can use the data sets that appear at the end of this investigation to fill in the cells in the table below.

	Home on Time	Not Home on Time	Total
Milwaukee Sample			60
Sweden Sample			56
Total			116

Looking at the Milwaukee data set, we can count the number of D answers in the question 5 column. Verify that this number is 19 and enter it into the table above. Use the data sets provided to fill in the remaining cells and compare your numbers to those in the following table.

	Home on Time	Not Home on Time	Total
Milwaukee Sample	19	40	60
Sweden Sample	14	42	56
Total	33	83	116

2. We can now compute the relevant sample proportions. The proportion who report that they would probably be home on time for the Milwaukee sample is 19/60 = .32. Verify that the corresponding proportion for the Sweden sample is .25.

3. Notice that the two sample proportions are different: .32 for the Milwaukee sample and .25 for the Sweden sample. The difference between the two sample proportions is .07. How can we tell if this difference is significant? We know that even if there was no difference in the true proportions for the two schools, we would still not expect the sample proportions to be exactly equal because of sampling variability (those chance differences that occur from one sample to another as a result of the random selection process). To determine if a difference of .07 is significant, we need to know something about what kinds of differences are consistent with sampling variability alone.

As you might have guessed based on the previous two investigations, we will carry out a simulation. In this simulation, we create a population consisting of 116 individuals (representing the 116 survey respondents). Do this by cutting out 116 squares of paper that are the same size. You can use the templates at the end of this investigation if you don't want to make your own. Mark 33 of the 116 squares with an "H" to represent the 33 survey participants who reported they would probably be home on time. The other 83 slips of paper, which are unmarked, will represent the 83 survey participants who reported that they would probably not be home on time.

Place all the slips of paper in a paper bag and mix them well. Next, select 60 slips of paper from the bag to represent the 60 Milwaukee students. Count the number of the 60 selected squares that have an H and enter the count into the highlighted cell in the table below.

	Home on Time	Not Home on Time	**Total**
Milwaukee Sample			**60**
Sweden Sample			**56**
Total	**33**	**83**	**116**

Use the row and column totals to fill in the remaining cells in the table. Then, compute the two sample proportions and the difference between the two sample proportions:

Home on Time proportion of Milwaukee sample =

Home on Time proportion of Sweden sample =

Difference in sample proportions (Milwaukee – Sweden) =

> *A difference is significant if it is larger than expected by chance alone.*

For example, if the number in the highlighted cell was 16, the resulting table, proportions, and difference would be the following:

	Home on Time	Not Home on Time	Total
Milwaukee Sample	16	44	60
Sweden Sample	17	39	56
Total	33	83	116

Home on Time proportion of Milwaukee sample = 16/60 = .27

Home on Time proportion of Sweden sample = 17/56 = .30

Difference in sample proportions (Milwaukee – Sweden) = .27 – .30 = – .03

4. Record the difference you obtained in step 3 in the Trial 1 row of the table below.

Trial Number	Difference in Sample Proportions
1	
2	
3	
4	
5	

5. Put all of the slips of paper back in the bag, mix well, and then repeat step 3 four more times. Use the templates below to keep track of your results and to compute the relevant proportions. Finally, enter the resulting differences in sample proportions into the trials 2 through 5 rows of the table in step 4.

Trial 2			
	Home on Time	Not Home on Time	Total
Milwaukee Sample			60
Sweden Sample			56
Total	33	83	116

Home on Time proportion of Milwaukee sample =

Home on Time proportion of Sweden sample =

Difference in sample proportions (Milwaukee – Sweden) =

Trial 3			
	Home on Time	Not Home on Time	Total
Milwaukee Sample			60
Sweden Sample			56
Total	33	83	116

Home on Time proportion of Milwaukee sample =

Home on Time proportion of Sweden sample =

Difference in sample proportions (Milwaukee – Sweden) =

Trial 4			
	Home on Time	Not Home on Time	Total
Milwaukee Sample			60
Sweden Sample			56
Total	33	83	116

Home on Time proportion of Milwaukee sample =

Home on Time proportion of Sweden sample =

Difference in sample proportions (Milwaukee – Sweden) =

Trial 5			
	Home on Time	Not Home on Time	Total
Milwaukee Sample			60
Sweden Sample			56
Total	33	83	116

Home on Time proportion of Milwaukee sample =

Home on Time proportion of Sweden sample =

Difference in sample proportions (Milwaukee – Sweden) =

6. OK, we are almost there! Each student in your class should have carried out five trials. Now combine all the simulated differences into one large data set by recording the values obtained by each student in your class in the table on the next page.

7. Use the data from step 6 to construct a dotplot of the simulated differences.

---

8. The dotplot of the simulated differences in step 7 provides information about how large the difference in sample proportions can be expected to be just as a result of sampling variability. That is, these simulated sample differences are consistent with the situation where there is no difference in the population proportions.

Now, let's go back and look at the actual difference in sample proportions for the Milwaukee and Sweden samples. The observed difference was .07. Based on what you see in the dotplot of simulated differences, is sampling variability a plausible explanation for why we might see a difference in sample proportions as large as .07? Explain why or why not.

9. Write a few sentences that address the original question posed: Is there a significant difference in the proportion of students who report that they would probably get home on time when given a curfew of midnight for students at the two schools?

10. In the last part of this investigation, your teacher will assign you to a team. Each team will either choose one of the questions below or formulate a question of its own that can be answered using the Milwaukee and Sweden data. Once your teacher has approved the team's question, use the data provided to answer it and write a brief report summarizing your results. (You should use a process similar to the one used to answer the "home on time" question, but note that because only the members of your team will be carrying out the simulation, you will each need to carry out more than five trials.)

Some possible questions to investigate:

Is there a significant difference in the proportion of students who report that they would give back an extra $10 in change for students at the two schools?

Is there a significant difference in the proportion of students who report that they would plan ahead and get a paper done without rushing and on time for students at the two schools?

Is there a significant difference in the proportion of students who report that they do homework right away for students at the two schools?

Data Set: Milwaukee							
Student #	Question (1)	Question (2)	Question (3)	Question (4)	Question (5)	Question (6)	Question (7)
1	Female	A	B	C	C	A	B
2	Female	D	D	D	D	B	D
3	Female	C	D	D	C	A	C
4	Female	A	D	D	D	C	C
5	Female	C	D	D	C	A	D
6	Female	C	D	D	D	D	C
7	Female	C	D	C	C	B	C
8	Female	D	D	C	C	C	C
9	Female	B	C	C	C	A	C
10	Female	D	C	C	B	A	C
11	Female	C	D	C	C	C	D
12	Female	C	B	C	C	D	C
13	Female	D	D	D	D	D	D
14	Male	A	D	C	C	C	B
15	Female	D	D	D	D	D	C
16	Male	D	B	C	D	D	C
17	Male	C	D	C	C	D	C
18	Male	D	D	D	D	D	D
19	Male	C	D	C	B	C	C
20	Male	C	B	D	C	A	C
21	Female	A	C	B	C	A	C
22	Male	C	D	D	B	D	C
23	Male	C	D	D	B	D	C
24	Male	C	D	A	A	B	C
25	Male	C	D	C	D	A	D
26	Female	C	D	D	C	D	C
27	Female	C	D	C	C	C	C
28	Male	C	D	A	C	D	C
29	Male	A	D	B	C	B	C
30	Female	C	C	C	B	C	C

Student #	Question (1)	Question (2)	Question (3)	Question (4)	Question (5)	Question (6)	Question (7)
				Data Set: Milwaukee			
31	Male	D	D	D	C	C	D
32	Male	B	B	C	B	D	D
33	Female	C	C	B	A	A	D
34	Female	C	D	C	C	A	C
35	Male	C	D	D	C	D	C
36	Male	B	C	C	D	A	C
37	Male	C	D	D	A	D	C
38	Female	D	D	D	D	D	D
39	Female	C	C	D	C	A	C
40	Male	C	D	B	C	D	C
41	Male	C	D	A	C	A	D
42	Male	C	B	C	C	D	C
43	Female	C	D	D	D	D	C
44	Male	D	D	C	C	B	D
45	Male	A	A	A	A	A	C
46	Female	C	D	C	D	D	C
47	Female	C	D	C	D	C	C
48	Male	D	C	A	D	D	C
49	Male	A	B	B	B	A	C
50	Female	C	D	B	B	A	C
51	Female	D	D	B	D	B	C
52	Male	B	C	B	A	A	C
53	Male	A	D	B	D	A	C
54	Female	D	D	C	C	C	D
55	Male	C	D	C	D	A	D
56	Female	A	D	D	D	C	C
57	Male	C	D	C	C	A	C
58	Female	C	C	B	C	C	C
59	Male	C	D	D	C	B	D
60	Female	C	D	D	D	D	C

Student #	Question (1)	Question (2)	Question (3)	Question (4)	Question (5)	Question (6)	Question (7)
				Data Set: Sweden			
1	Female	C	D	C	B	D	C
2	Female	C	D	C	B	D	C
3	Female	B	D	B	B	D	C
4	Male	C	C	B	B	A	C
5	Male	C	B	D	D	A	C
6	Female	C	C	B	C	A	C
7	Female	C	B	B	B	B	C
8	Male	A	B	D	C		C
9	Female	D	D	D	C	C	D
10	Female	C	C	D	C	B	C
11	Female	C	D	D	D	D	C
12	Male	B	D	D	D	B	C
13	Female	C	D	B	D	D	D
14	Female	D	C	B	C	D	D
15	Female	D	C	C	C	A	D
16	Male	D	C	D	C	C	C
17	Male	C	D	D	D	A	D
18	Male	C	C	C	C	A	C
19	Female	C	D	C	C	D	C
20	Female	D	C	D	C	A	D
21	Female	D	D	C	D	A	C
22	Male	C	B	C	C	A	C
23	Female	C	D	D	C	D	C
24	Female	B	C	C	B	A	C
25	Male	D	C	C	D	A	C
26	Female	C	D	D	C	D	D
27	Female	C	D	D	C	A	C
28	Female	D	D	C	D	D	C

Data Set: Sweden							
Student #	Question (1)	Question (2)	Question (3)	Question (4)	Question (5)	Question (6)	Question (7)
29	Female	D	D	D	D	C	C
30	Female	D	D	C	C	C	C
31	Male	B	D	C	C	D	D
32	Female	C	D	B	D	A	C
33	Female	C	D	B	C		C
34	Male	C	D	D	C	A	B
35	Female	C	D	D	C	D	C
36	Female	B	D	D	C	B	D
37	Female	C	D	D	C	B	C
38	Male	C	C	C	B	D	C
39	Female	C	C	B	C	A	C
40	Female	C	C		C	D	C
41	Female	C	C	D	C	B	D
42	Female	B	C	D	C	A	B
43	Male	D	C	C	C	A	D
44	Female	C	D	C	D	B	D
45	Female	C	C	C	C	B	C
46	Male	C	D	D	C	A	C
47	Male	C	D	D	C	A	C
48	Female	C	C	D	D	A	D
49	Male	C	C	C	B	A	C
50	Female	D	D	C	C	D	D
51	Male	C	D	D	D	D	D
52	Male	C	D	D	D	D	C
53	Male	B	C	D	C	A	C
54	Male	C	D	D	C	C	C
55	Female	D	D	D	B	B	D
56	Female	C	C	D	C	D	C