Hiring discrimination—it just won't fly!

An airline has just finished training 25 pilots—15 male and 10 female—to become captains. Unfortunately, only eight captain positions are available right now. Airline managers announce that they will use a lottery to determine which pilots will fill the available positions. The names of all 25 pilots will be written on identical slips of paper, placed in a hat, mixed thoroughly, and drawn out one at a time until all eight captains have been identified.

A day later, managers announce the results of the lottery. Of the 8 captains chosen, 5 are female and 3 are male. Some of the male pilots who weren't selected suspect that the lottery was not carried out fairly. Do these results provide *convincing* evidence of discrimination?

What is the evidence of discrimination?

What are the two explanations for this evidence?

How can we determine if the evidence of discrimination is convincing?

What's the difference between a population and a sample? What is a census?

Sampling Badly

What's the problem with convenience samples?

What is bias?

What's a voluntary response sample?

Example:
An AP® Statistics teacher was curious about the average grade point average (GPA) of students at his school. He used the 32 students in his second-period AP® Statistics class as a sample and concluded that the average GPA of students at his school is about 3.85.

What type of sampling did the teacher use?

Explain how this plan will result in bias and how the bias will affect the estimated proportion.

Good Sampling
Simple Random Sample (SRS)

The manager of a mall plans to survey a random sample of 3 stores to determine the hours they would like to stay open during the holiday season.

Aeropostale	Forever 21	Old Navy
All American Burger	GameStop	Pac Sun
Arby's	Gymboree	Panda Express
Barnes & Noble	Haggar	Payless Shoes
Carter's for Kids	Just Sports	Star Jewelers
Destination Tan	Mrs. Fields	Vitamin World
Famous Footwear	Nike Factory Store	Zales Diamond Store

Use your calculator to select an SRS of size 3 stores.

Use the table of random digits below to select an SRS of size 3 stores.

LINE								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056

What's the difference between sampling *with* replacement and sampling *without* replacement? How should you account for this difference when using a table of random digits or other random number generator?

Stratified Random Sample

How is it different than a simple random sample?

Cluster Sample

How is it different than a stratified sample?

Systematic Random Sample

A librarian wants to know the mean number of pages in all the books in the library. The library has 20,000 books, arranged by type (fiction, biography, and so on) in shelves that hold about 50 books each.
(a) Explain how to select a simple random sample of 500 books

(b) Explain how to select a stratified random sample of 500 books. Explain your choice of strata.

(c) Explain how to select a cluster sample of 500 books. Explain your choice of cluster.

(d) Explain how to select a systematic random sample of 500 books.

(e) Which method do you prefer? Why?

What Can Go Wrong

Undercoverage

Nonresponse

Response Bias

Question Wording Bias

1. Each of the following is a possible source of bias in a sample survey. Name the type of bias that could result.

a. The sample is chosen at random from a school directory.

b. Some people cannot be contacted in five calls.

c. Interviewers choose people walking by on the sidewalk to interview.

2. A survey paid for by makers of disposable diapers found that 84% of the sample **opposed banning** disposable diapers.

Here is the actual question: "It is estimated that disposable diapers account for less than 2% of the trash in today's landfills. In contrast, beverage containers, third-class mail, and yard wastes are estimated to account for about 21% of the trash in landfills. Given this, in your opinion, would it be fair to ban disposable diapers?"

Do you think the estimate of 84% is less than, greater than, or about equal to the percent of all people in the population who would oppose banning disposable diapers? Explain your reasoning

Experiments

Last year RHS offered an after-school SAT prep class that students could volunteer to take. 44 students took the course and then took the SAT. The average SAT score for this group was 1220. The average SAT score for all students who did not take the prep class was 1050.

Is the situation described an observational study or an experiment?

Identify the explanatory variable and the response variable.

Can you conclude that taking the prep course will cause a student's SAT score to increase? Why or why not?

Identify as many other possible variables that you can that may explain why the SAT scores are higher for those who took the prep course than for those who did not.

Design an experiment that would allow us to determine if the SAT prep causes an increase in SAT scores. Be as thorough as possible.

Does reducing screen brightness increase battery life in laptop computers? To find out, researchers obtained 30 new laptops of the same brand. They chose 15 of the computers at random and adjusted their screens to the brightest setting. The other 15 laptop screens were left at the default setting— moderate brightness. Researchers then measured how long each machine's battery lasted. Was this an observational study or an experiment? Justify your answer.

What is the difference between a prospective and a retrospective observational study?

Define the following terms:

- Treatment

- Experimental unit/subject

- Factor

- Level

List the 4 principles of experimental design

Would you fall for the placebo effect? Let's watch a video

What is the placebo effect?

Similar to the video, we want to use a beverage to test the affect that caffeine can have on heart rate. Here is an initial plan:

measure initial pulse rate
give each student some caffeine (Coca-Cola)
wait for a specified time
measure final pulse rate
compare final and initial rates

What are some problems with this plan? What other variables will be sources of variability in pulse rates?

What is the purpose of a control group?

What is a single-blind experiment? A double-blind experiment? Why is blindness important?

What is the purpose of random assignment?

How do we randomly assign treatments to units (or units to treatments)?

What is replication?

Back to the caffeine experiment

We change the experiment so that students are randomly assigned to drink Coke with caffeine and Coke with out caffeine. We compare the pulse rates of the two groups.

Based on the possible results of the caffeine experiment, what is the evidence that one of the drinks increases pulse rates more than the other?

What are the two explanations for this evidence?

How can we determine if the evidence is *convincing*?

What does it mean if the results of an experiment are statistically significant?

Many utility companies have introduced programs to encourage energy conservation among their customers. An electric company considers placing small digital displays in households to show current electricity use and what the cost would be if this use continued for a month. Will the displays reduce electricity use? One cheaper approach is to give customers a chart and information about monitoring their electricity use from their outside meter. Would this method work almost as well? The company decides to conduct an experiment using 60 households to compare these two approaches (display, chart) with a group of customers who receive information about energy consumption but no help in monitoring electricity use.

Explain why it was important to have a control group that didn't get the display or the chart.

Create an outline showing a completely randomized design for the experiment. Make sure to explain how you randomly assign treatments.

Back to the SATs

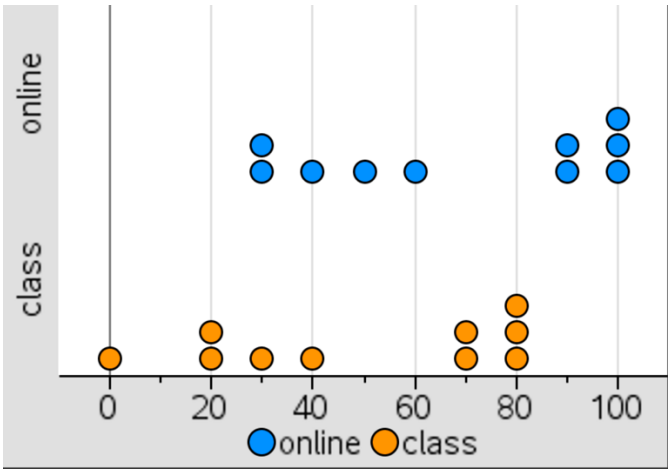
Many students enroll in prep courses to improve their SAT scores. Twenty students who have taken the SAT once volunteered to participate in an experiment comparing online and classroom prep courses.

Describe how we can use a completely randomized design to compare online and classroom SAT prep courses.

Among the 20 volunteers, 10 are in Precalculus, 6 are in Algebra 2, and 4 are in Geometry. What problem does this cause?

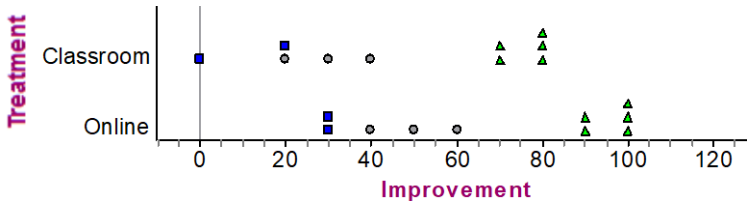
How can we address this problem?

Here are the results of the experiment, using math level as a blocking variable. Based on the dotplots, does there appear to be convincing evidence that the online course is better?



Class	Treatment	Improvement
P	Online	100
P	Online	100
P	Online	90
P	Online	90
P	Online	100
P	Classroom	70
P	Classroom	70
P	Classroom	80
P	Classroom	80
P	Classroom	80
A	Online	50
A	Online	60
A	Online	40
A	Classroom	30
A	Classroom	40
A	Classroom	20
G	Online	30
G	Online	30
G	Classroom	0
G	Classroom	20

The dotplots above ignored the fact that we blocked by math level. Here is the dotplot again, using different symbols for students in each math level.



Notice that within each math level, the online students clearly did better. We couldn't see this difference when we ignored the blocks because of the variability created by math class.

What is the benefit of blocking?

In general, how can we determine which variables might be best for blocking?

What are some variables that we can block for in the caffeine experiment?

What is the difference between blocking and stratifying?

What is a matched pairs design?

What is inference?

What is sampling variability?

What is a margin of error?

Inference about the population versus cause & effect.

		Were individuals randomly assigned to groups?	
		Yes	No
Were individuals randomly selected?	Yes	Inference about the population: YES Inference about cause and effect: YES	Inference about the population: YES Inference about cause and effect: NO
	No	Inference about the population: NO Inference about cause and effect: YES	Inference about the population: NO Inference about cause and effect: NO

Many students insist that they study better when listening to music. A teacher doubts this claim and suspects that listening to music actually hurts academic performance. Here are four possible study designs to address this question at your school. In each case, the response variable will be the students' GPA at the end of the semester.

1. Get all the students in your AP Statistics class to participate in a study. Ask them whether or not they study with music on and divide them into two groups based on their answer to this question.
2. Select a random sample of students from your school to participate in a study. Ask them whether or not they study with music on and divide them into two groups based on their answer to this question.
3. Get all the students in your AP Statistics class to participate in a study. Randomly assign half of the students to listen to music while studying for the entire semester and have the remaining half abstain from listening to music while studying.
4. Select a random sample of students from your school to participate in a study. Randomly assign half of the students to listen to music while studying for the entire semester and have the remaining half abstain from listening to music while studying.

For each design, suppose that the mean GPA for students who listen to music while studying was significantly lower than the mean GPA of students who didn't listen to music while studying. What can we conclude for each design?