Know the difference between an experiment and an observational study, (13, 5a; 11b, 2a; 07, 5a; 03b, 3a; M2, 1)

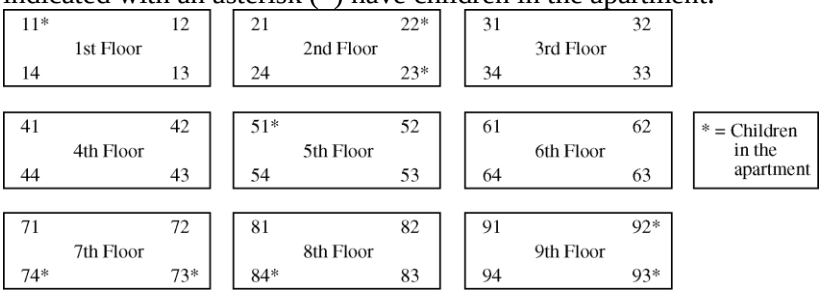
In an experiment someone is implementing change in the environment. If I were testing the relationship between a special diet and heart disease, I would take 1000 people and randomly make half do a special diet, while the other half would continue to do what they normally do.

In an observational study there is no implementation of change. We are simply observing people in their natural habitat. A study on the relationship between diet and heart disease would look at 1000 people that are in their 70s. We would look at those who had "good" 3diets and those who had "bad" diets and relate that data to whether or not they have had heart problems.

Observations Studies and Collecting Data

An apartment building has nine floors and each floor has four apartments. The building owner wants to install new carpeting in eight apartments to see how well it wears before she decides whether to replace the carpet in the entire building.

The figure below shows the floors of apartments in the building with their apartment numbers. Only the nine apartments indicated with an asterisk (*) have children in the apartment.



Know how to create a cluster sample, (11, 3a)

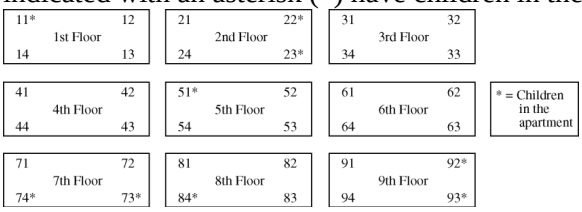
A cluster sample is where you choose to survey only parts of a population. So, if I am running for governor, a cluster sample would be only going to 3 out of the 10 counties. This is not the same as a stratified sample. In a stratified sample, you are guaranteeing that every part (or strata) of a population is being represents. So, if I am running for governor, I would make sure to take a sample from every county. That way every county is represented.

Let's create a cluster sample for the apartment problem above. We will make sure to incorporate the design and randomization:

> "In this case, I would call each floor a cluster. So, one way to do a cluster sample would be to pick 2 floors at random and then carpet every apartment on those two floors. That would give us the 8 apartments that we want. For randomization, I will use a random number generator to generate two unique numbers between 1 and 9. We will then carpet the two floors corresponding to the two numbers generated."

An apartment building has nine floors and each floor has four apartments. The building owner wants to install new carpeting in eight apartments to see how well it wears before she decides whether to replace the carpet in the entire building.

The figure below shows the floors of apartments in the building with their apartment numbers. Only the nine apartments indicated with an asterisk (*) have children in the apartment.



Know what a stratified random sample is, how to describe it so a reasonable person could do it, and how to implement one where you need to represent different proportion in a population (10% Hispanic, 20% Native American, etc), (13, 2c; 10, 4c;10b, 2b, 05, 5c; M7, 20; M2, 15)

Let's create a stratified sample for the apartment problem above. We will make sure to describe why we chose the variable that we did, incorporate the design, and describe the randomization.

"In this case, I would choose to stratify by whether or not the apartments have children. I chose this variable, because I believe that it will have the biggest impact on our response variable (carpet durability). Since 25% of the apartments have children, I will make sure that 25% of my sample (2 of 8) will be apartments with children. I would first start with the seven apartments that have children. I would write the apartment numbers corresponding to those apartments on nine small, equal-sized pieces of paper. I will mix up those nine papers in a hat and then randomly choose 2 of them. The apartments corresponding to those two apartments will receive carpeting. I will then do a similar process for the 27 apartments without children, and I will chose 6 apartments at random."

Know how to take a simple random sample from a large population in multiply ways, and how to describe it so a reasonable person could do it, (14, 4b; 13, 2b; 12, 6a; 10b, 2a;08, 2c; 04b, 2a)

Using the apartment problem above, we could take a SRS from the population of 36 apartments using a random number generator. To do this, I would first assign each apartment a number from 1 to 36. Then, I would use a random number generator to generate 8 unique numbers between 1 and 36. The apartments that correspond to those numbers will be the apartments used for our sample.

Why is a stratified sample sometime better than a SRS? Better than a cluster? (11, 3b; 10b, 2c)

A stratified sample is "better" because it guarantees that certain parts of the population are represented. For the apartment problem above, a stratified sample would be better than a SRS or cluster sample, because it guarantees that an apartment with children will be chosen. If a SRS or cluster sample were done, there is a chance that 8 apartments (in the SRS) or 2 floors (in the cluster) could be chosen that do not have any apartments with children. This would be bad, because children could definitely have an impact on the durability of carpet.

Know what a census is, (M7, 2)

A census is when every individual in a population is represented. In the apartment problem above, a census would be when every apartment gets the new carpeting. A census is usually incredibly difficult to get. For example, it would take a lot of time and energy to get a census at CV. Here the setback of a census would be that it is costly.

Know that you can only make conclusions about the population that your sample is taken from (11b, 2b; 06, 5cd; 05, 1b; 04, 3c; 04, 5b; 03, 4c; 03b, 4d, M7, 16)

If I choose five people from 4th period I can make conclusions about 4th period. If I take 5 people from CV, I can make conclusions about CV. I cannot use either of those samples to draw conclusions about CV, because my sample did not come from CV. It came from subsets within CV.

Know what bias is and be able to explain in context how a bias can directly affect the results (the proportion would be higher if the sample was truly random, (14, 4b; 13, 2a; 08, 2a; 05, 5a; 04b, 2a; M7, 9)

Bias does not always occur on purpose. Let's say a poll is taken in the quad at lunch about whether or not students should have to pay in order to take advantage of the open campus policy at lunch. Below is how we might comment on bias. Notice, that we say how we think this bias will affect the outcome (higher estimate). This should always be done when talking about bias.

"Bias is introduced into this survey by when the survey is taken. Since students in this sample are on campus for lunch, they might be indifferent to what the school lunch policy is. I believe that this will lead to higher estimate for the proportion of students who think students should have to pay. If people who were already off campus for lunch were surveyed, they would be much more likely to answer that paying should not be necessary. Therefore, I think this survey will result in an overestimate of the population of students who think paying for open campus is appropriate."

Know what wording bias is, how to fix it, and how it affects an outcome in the context of the problem, (04b, 2b)

A survey taken at CV asked, "It was determined that La Canada's prom cost $30 less per student than CV's prom. Do you think that CV charges too much for prom?" How is bias introduced in this question.

Bias was introduced by the first sentence, which makes a reference to CV's prom. I believe that this is leading information that is trying to persuade the reader to answer yes. I believe that this could lead to an overestimate of the population who are concerned about prom tickets. If that sentence were removed, we would most likely get a beter estimate of people who think that prom is over priced. To fix this problem I would simply ask, "Do you think that CV charges too much for prom?"

Note: We again make sure to note how the bias will affect our data (overestimate).

Experimental Design

Agricultural experts are trying to develop a bird deterrent to reduce costly damage to crops in the United States. An experiment is to be conducted using garlic oil to study its effectiveness as a nontoxic, environmentally safe bird repellant. The experiment will use European starlings, a bird species that causes considerable damage annually to the corn crop in the United States. Food granules made from corn are to be infused with garlic oil in each of five concentrations of garlic —0 percent, 2 percent, 10 percent, 25 percent, and 50 percent. The researchers will determine the adverse reaction of the birds to the repellant by measuring the number of food granules consumed during a two-hour period following overnight food deprivation. There are forty birds available for the experiment, and the researchers will use eight birds for each concentration of garlic. Each bird will be kept in a separate cage and provided with the same number of food granules.

Know what a treatment is and how to list them based on the description of an experiment with more than one factor? (10, 1a; 06, 5a; 06b, 5a)

An experiment is when a treatment is being imposed on an individual experimental unit. In the garlic oil question above there are 5 treatments: 0%, 2%, 10%, 25%, 50%.

What is an experimental unit? (10, 1a; 06b, 5a)

An experimental unit is what is getting the treatment. In the garlic oil problem above, the experimental units would be the forty starlings. Each starling is receiving one of the 5 treatments.

What is a response variable? (10, 1a; 06b, 5a)

The response variable is what we are measuring from the experimental units. In the garlic oil problem above, the response variable is the number of food granules consumed. The response variable is what we would use to run our test. In this case, we would probably run a Chi Squared Goodness of Fit Test.

Before beginning a unit on frog anatomy, a seventh-grade biology teacher gives each of the 24 students in the class a pretest to assess their knowledge of frog anatomy. The teacher wants to compare the effectiveness of an instructional program in which students physically dissect frogs with the effectiveness of a different program in which students use computer software that only simulates the dissection of a frog. After completing one of the two programs, students will be given a posttest to assess their knowledge of frog anatomy. The teacher will then analyze the changes in the test scores (score on posttest minus score on pretest).

Describe a method for assigning the 24 students to two groups of equal size that allows for a statistically valid comparison of the two instructional programs.

Why wouldn't it be appropriate to assign people by flipping a coin at times? (09, 3a)

One students said the following

> "I will first put the students in alphabetical order. Then, I will call up each students one by one in that order. I am going to flip a coin for each students. If the coin is heads, then they will physically dissect the frog. If it is tails, they will use the computer software. I will continue this until one group contains twelve, and the remaining students will be placed in the other group. After, we will compare the changes in test scores"

This seems like an appropriate way to randomize, except in the case where the students with the last names at the end of the alphabet have similar abilities. Let's say that the last five students alphabetically have the five lowest grades in the class. If the computer software group fills up after the 18^{th} students, then the 5 students with the lowest grades are going to go in the physical dissection group. This could cause bias in the results, because these students have much more room for growth. So, the teacher could conclude that physical dissection is better, but we only got those results because of bad randomization.

It's not bad to flip a coin or roll a die for randomization, you just have to make sure that students are being chosen in an order that won't bias the results in a way that it was above. I would probably choose to do something like what is written in the next sections.

Know how to create a completely randomized design experiment, including how to randomly assign your experimental units so that a reasonable statistician could do it, (09, 3a; 08b, 4a; 07, 2b; 06, 5b; 06b, 5b; 05b, 3a; 03b, 4a; M7, 35; M2, 25)

For the frog anatomy question above you want to make sure that your response includes 3 things

1) Design
2) Randomization to a specific treatment
3) Response Variable

So, we could write something like this:

> "I use a random number generator to assign every students a unique number between 1 and 24. The students with number 1 through 12 will do a physical dissection, while the students who were assigned 13-24 will use the computer software. After they have finished the programs, we will then compare their scores on the posttest and pretest."

Note: On a side note, if we wanted to test for significance in this situation, we would do a paired t-test. This is because we would be looking on how much students improved (post – pre). We don't really care what they got on the individual tests, just the difference.

Before beginning a unit on frog anatomy, a seventh-grade biology teacher gives each of the 24 students in the class a pretest to assess their knowledge of frog anatomy. The teacher wants to compare the effectiveness of an instructional program in which students physically dissect frogs with the effectiveness of a different program in which students use computer software that only simulates the dissection of a frog. After completing one of the two programs, students will be given a posttest to assess their knowledge of frog anatomy. The teacher will then analyze the changes in the test scores (score on posttest minus score on pretest).

Describe a method for assigning the 24 students to two groups of equal size that allows for a statistically valid comparison of the two instructional programs.

Why is it important to assign individuals in an experiment as opposed to letting them pick? (11b, 2c; 09, 3b; 03, 4a)

For the frog anatomy problem above, it may not be a good idea to let students pick what groups they are in, even if you guarantee that each group has twelve. The problem is if certain groups who are prone to a certain result put themselves in a group. Let's say for example that there is a class of 12 boys and 12 girls. In this particular class, for whatever reason, the 12 girls knew a great deal about frogs before this experiment. They would, therefore, be more likely to do well on the pretest. This would make it harder to do better on the posttest because they did so well on the pretest. Let's also say that the girls all want to stay together and they do not want to take part in the dissection. So, all the girls chose the computer software. In this case, the teacher may conclude that the dissection is the better method, because the boys showed a much bigger difference in their test scores. The problem is that this was not caused by the learning process, it was caused by the fact that the girls wanted to do the computer program, and they knew a lot about frogs to begin with.

Know what confounding is and how it can affect the results of an experiment, (06b, 5d)

The basic confounding problem is the firefighters versus fire damage study. When looking at the number of firefighters and the amount of damage caused by a fire, we might conclude that more fire fighters means more damage. So, should we stop sending so many firefighters? Obviously, no. Confounding has occurred, because we are not accounting for a third variable, the size of the fire.

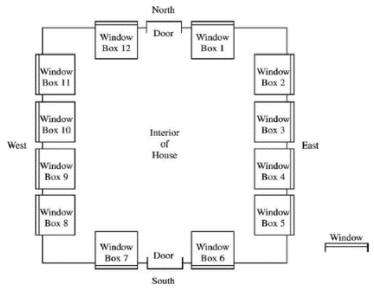
When talking about confounding, you want to do two things

1) Say what is causing the confounding
2) Say how the confounding causes us to be unable to determine which variable is causing the response.

Let's use the answer given in the previous section about the students picking what group they want to be in for the frog experiment. Here is how we could talking about confounding in this situation

> "Confounding is when we can't determine what variable is causing the response. In the above situation, we would presume that the boys would show a great gain in their test results. So, confounding is caused by the fact that student got to pick which group they were in. In this case, confounding occurs because we are unable to determine if it was the frog dissection or the self-selection process that caused the change in test scores."

The United States Department of Energy is conducting an experiment to compare the heat gain in houses using two different types of windows, A and B. Six windows of each type are available for the experiment. The Department has constructed a house with twelve windows as shown on the floor plan below.



In the interior of the house, each window is surrounded by a window box to capture and measure the amount of heat coming in through that window and to isolate the heat gain for each window.

Know what a block design is and how to randomly assign experimental units within that design, (09b, 4a; 07b, 3b; 04, 2ab; M2, 16)

For the heat gain problem above, we can create a block design for this experiment. The purpose of creating a block design is because we believe that groups of experimental units (windows in this case) will have a have a similar response related to the response variable. In this case, it would reason that windows on one side of the house will have similar measures of heat gain, regardless of the type of window. For example, windows on the west side may get stronger exposure to sun in the afternoon that any other side of the house. Therefore, those windows are going to have higher heat gain than the other windows because of where they. This could create confounding if a completely randomized design randomly assigned all of window A to the west side. This would cause us to think that window A allows a lot of heat in, when it is because they are on the west side. So, below is a description of how to block:

"Since different sides of the house could experience different amounts of heat (response variable), I would block by side of the house. So, in this case I would create four blocks: North (windows 1, 12), South (windows 6, 7), East (windows 2,3,4,5), and West (windows 8,9,10,11). For each block I would randomly assign half the windows boxes with window A, and the other half would get window B. For example, on the east side of the house I would place numbers 2,3,4, and 5 on equal sized pieces of paper and place them in a hat. After mixing them up, I would randomly pull out two pieces of paper. The window boxes corresponding to the pulled numbers would receive window A, while the remaining two window boxes would receive window B. I would do a similar process for the remaining three sides of the house. We would then compare the amount of heat coming through the different brands of window."

Know why it is beneficial to do a block design at times, and why it is important to create homogeneous groups, (07, 2c; 07b, 3a; 04, 2c; M7, 14; M7, 31)

In the window problem above a block design is better, because it REDUCES (Not eliminates) chances for confounding. As said in the previous example, if we "accidentally" randomly assigned window type A to the west side, we might think that window A is bad at keeping out heat. But, in reality those windows were exposed to more heat because they were on that side of the house. So, we put all of the window boxes on the west side into one block, because we think that they will have similar results because of where they are.

As dogs age, diminished joint and hip health may lead to joint pain and thus reduce a dog's activity level. Such a reduction in activity can lead to other health concerns. A study is to be conducted to see which of two dietary supplements, *glucosamine* or *chondroitin*, is more effective in promoting joint and hip health and reducing the onset of canine osteoarthritis. Researchers will randomly select a total of 300 dogs from ten different large veterinary practices around the country. All of the dogs are more than 6 years old and their owners have given consent to participate in the study. Changes in joint and hip health will be evaluated after 6 months of treatment.

Know what a control group is, why we control experiments, and how to describe its benefits in the context of a problem, (07, 2a; 03, 4b; 03b, 4b)

A control group gives us a "baseline" to compare our treatments to. In general a control group (sometimes given a placebo) lets us know how people in our experiment would react if they were given nothing. For the dog information above, students were asked to explain the purpose of a control group. Here is an example or a response:

> "In this case, we have two types of drugs, creating two treatment groups. It might be beneficial, though, to add a third group that will received neither of the treatments. This will be our control group. I don't think that we will need to give the third group a placebo, because there is no reason to believe that dogs will even know that they are taking a drug. The purpose of creating this third control group is to let us know whether or not the drugs do anything at all to affect hip and joint health. For example, we might find that both of the drugs actually were worse for hip and joint health than doing nothing would have been."

Know what it means for an experiment to be double blind and how it can be implemented, (09b, 6a)

Blinding an experiment makes it so people involved in the experiment are not aware of the treatment they are getting. Let's use the example of a drug that is meant to help reduce blood pressure. We might create a randomly designed experiment where half of the people get the drug and the other half get a placebo.

Single Blind: It would be beneficial to make the experiment single blind by not letting the subjects know what treatment they are receiving. We would not want an individual knowing that they are getting a placebo, because that could affect our results.

Double Blind: It might also be beneficial to make this experiment double blind by not telling the subjects what they are getting and by making the doctors unaware of which treatment they are prescribing. A doctor could introduce bias (purposefully or accidentally) if he/she knew which subjects got which treatments. If a doctor knew he/she was giving the treatment, he/she might put a higher initial blood pressure, so that the results would show a bigger drop for people getting the drug.

Know how replication is used to improve a study/experiment and how it is implemented correctly, (06b, 5c; 05, 1c)

Replication is a key component to any experiment, because it allows us to minimize the results of outliers that could occur randomly. Let's look at this experiment:

A study was conducted to determine whether a newly developed fuel would be able to increase the miles per gallon compared to an older fuel. Two different cars were used in this study. It was randomly determined which car would get the newly developed fuel. As the car drove, a measurement device on the car automatically recorded the amount of gasoline consumed at 25 randomly selected points. After the car finished its drive, the second was filled with the older fuel and drove the same route. Twenty-five measurements of gasoline consumption were also recorded at randomly selected points.

Just because we took 25 observation, doesn't mean that we replicated. We only used two cars. So, replication was not used correctly. Imagine trying to compare a blood pressure drug and only having two people. It wouldn't matter if we took 1000 measurement from those two people, we still only have two people. Replication has not been done.

Know how to create a matched pairs design, including how you randomly assign treatments, (05b, 3b; 04b, 4b)

There are two types of paired design:

Method 1: Have every subject do both treatments.

If we were testing if people liked Coke or Pepsi doing a blind test, we could have everyone blindly drink both and the have them give each one a score between 1 and 10. Since every experiment needs randomization, we would randomly assign which soda they drank first by flipping a coin. We are essentially creating several blocks of size 1.

Method 2: Pair two similar experimental units together and have each one do a different treatment.

If we were testing two fitness programs that improved 40 yard dash time, we could pair by initial speed. We would take all of our volunteers and have them run a 40 yard dash. Afterwards, we would take the two fastest volunteers and randomly assign one of them to one fitness program, and the other volunteer would do the other program. We would do this for the next two fastest individuals, and we would continue this process until we got to the two slowest volunteers. After the programs were complete, we would calculate 40 yard dash times again and compare them to determine if one of the programs is better.We are essentially creating several blocks of size 2.

Note: If we did a hypothesis test with the data collected in either method, it would be a paired t-test. Even though both experiment result in two sets of data, we would be more concerned with data created by the difference in the two results.