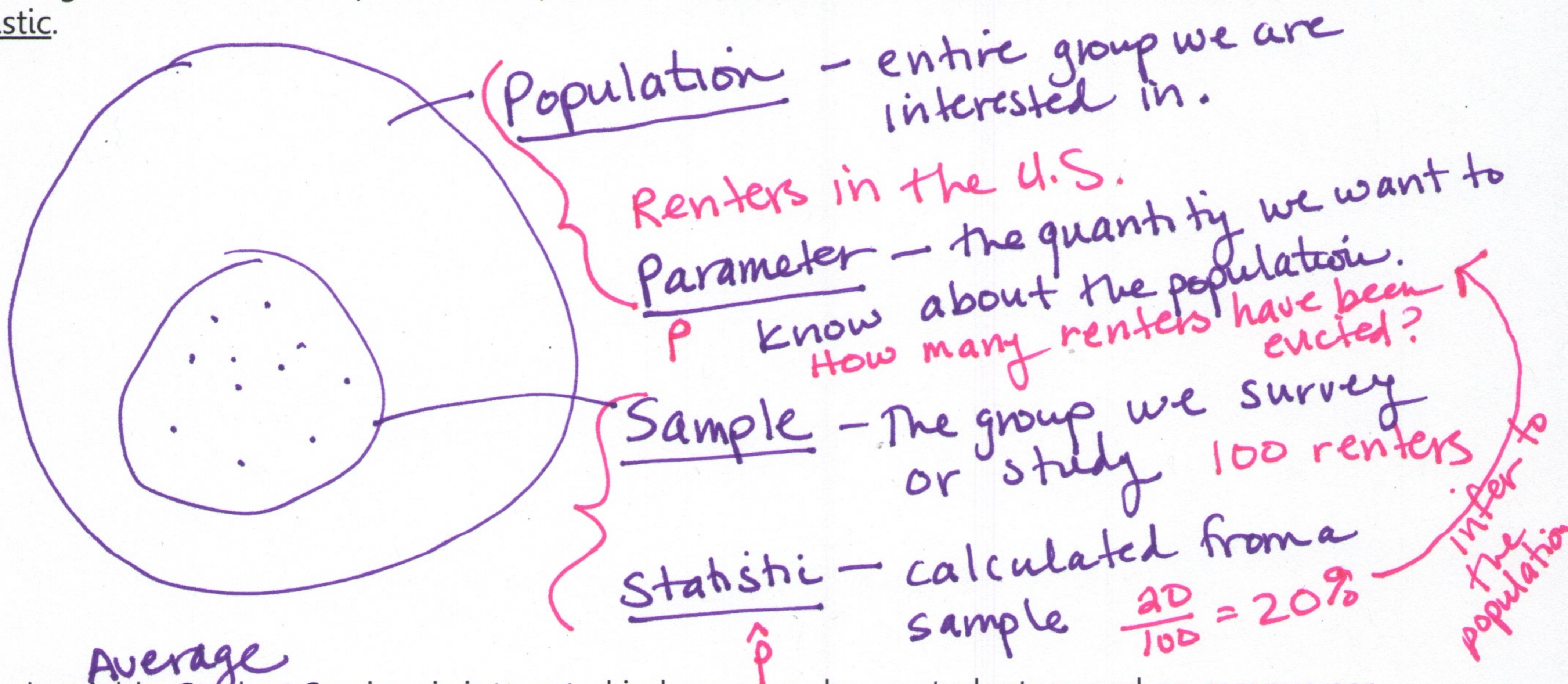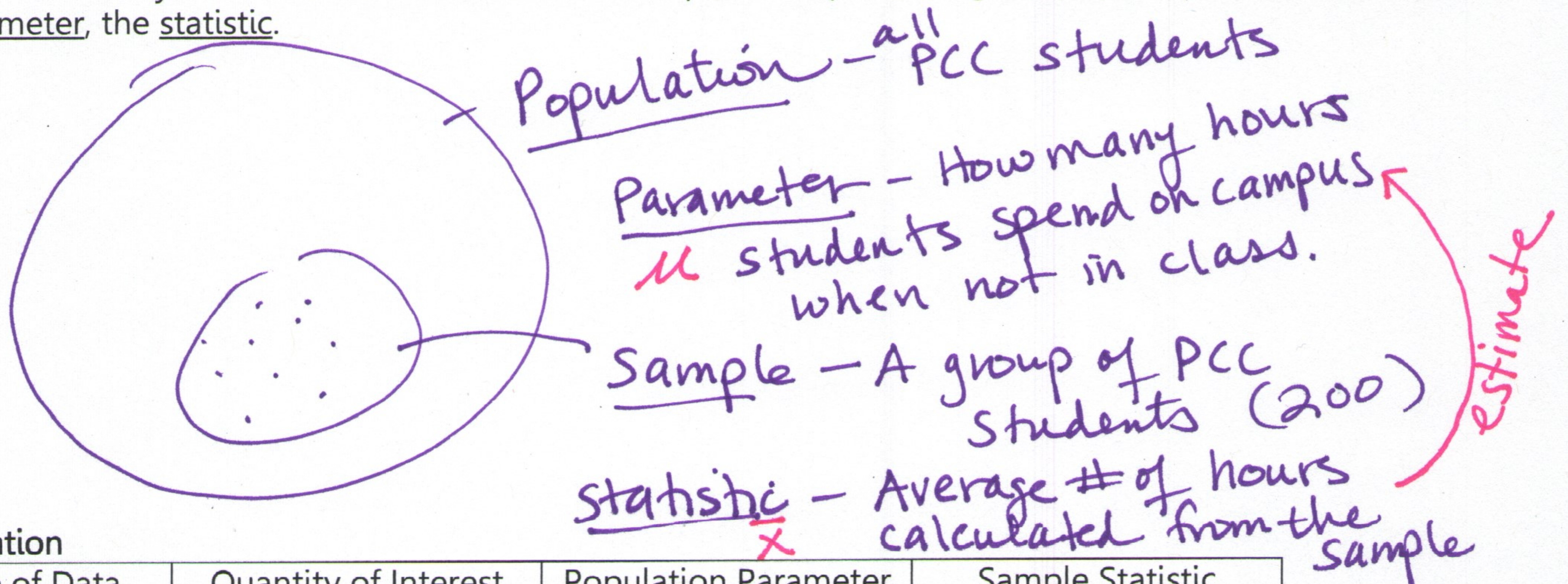Overview
- Population, Sample, Parameter and Statistic
- Anecdotal Evidence, Observational Studies, Experiments
- Sampling Methods and Types of Bias
- Experimental Design

Population, Sample, Parameter and Statistic

Categorical yes/no variable: *%* Researchers are interested in how many people have been evicted from *in the U.S.* their housing. Draw and label a picture to represent the population, the sample, the parameter, and the statistic.

Population – entire group we are interested in.

Renters in the U.S.

Parameter – the quantity we want to know about the population. p
How many renters have been evicted?

Sample – The group we survey or study   100 renters

Statistic – calculated from a sample $\hat{p}$ $\frac{20}{100} = 20\%$
infer to the population

Average
Numerical variable: Student Services is interested in how many hours students spend on campus per week when they are not in class. Draw and label a picture representing the population, the sample, the parameter, the statistic.

Population – *all* PCC students

Parameter – How many hours students spend on campus when not in class. μ

Sample – A group of PCC students (200)

Statistic – Average # of hours calculated from the sample $\bar{x}$ *estimate*

Notation

Type of Data	Quantity of Interest	Population Parameter	Sample Statistic
Categorical (yes/no)	Proportion	p	$\hat{p}$ p-hat
Numerical	Mean	μ mu	$\bar{x}$ x-bar

Anecdotal Evidence — not a randomized method
A single or small number of results, often unusual cases that are not representative of the population.

Example: *A person on the news got mercury poisoning from eating swordfish, so the average mercury concentration in swordfish must be dangerously high.*

Observational Studies
Surveys are also observational

In an observational study, researchers gather data without interacting with the subjects.

Retrospective	Prospective
Looking at past records to get the data	Subjects are identified and followed over time.

Randomized Experiments
In a controlled, randomized experiment, researchers assign treatments to groups of subjects and measure a response variable.

Match the description with the proper term.

a. A group of disabled women aged 65 and older were tracked for several years, ending in 2010. Those who had a vitamin B12 deficiency were found to be twice as likely to suffer severe depression as those who did not.

b. Researchers want to investigate whether taking aspirin regularly reduces the risk of heart attack. Four hundred people who identify as men are divided randomly into two groups: one group will take aspirin, and the other group will take a placebo. At the end of the study, researchers count the number of men in each group who have had heart attacks.

c. Researchers who examined health records of thousands of males found that men who died of myocardial infarction (heart attack) tended to be shorter than men who did not.

d. A doctor worked with two patients whose depression was cured with vitamin B12 injections.

1. Retrospective Observational Study

2. Prospective Observational Study

3. Anecdotal Evidence

4. Experiment

e. In part c above, is it correct to conclude that shorter men are at higher risk of dying from a heart attack? Could there be a **lurking or confounding variable**?

No, there must be a lurking or hidden variable. genetics, diet, environment, health history, exercise, age

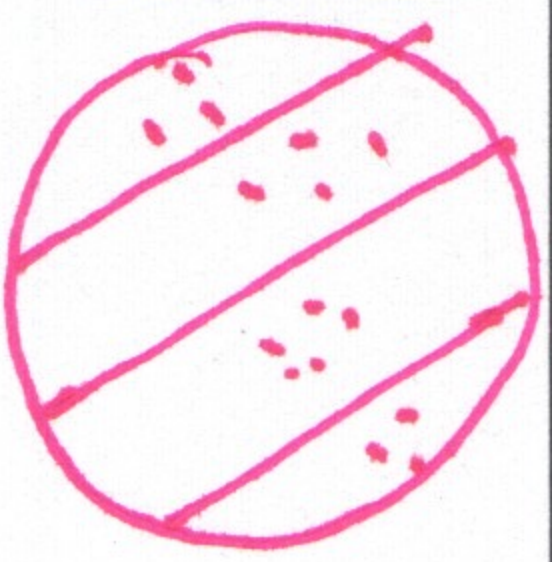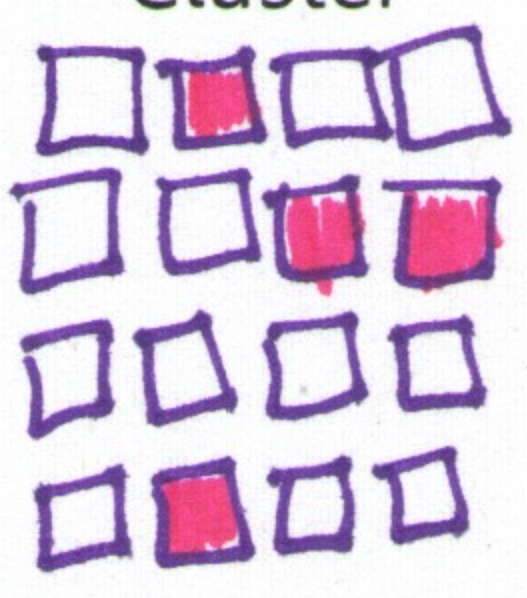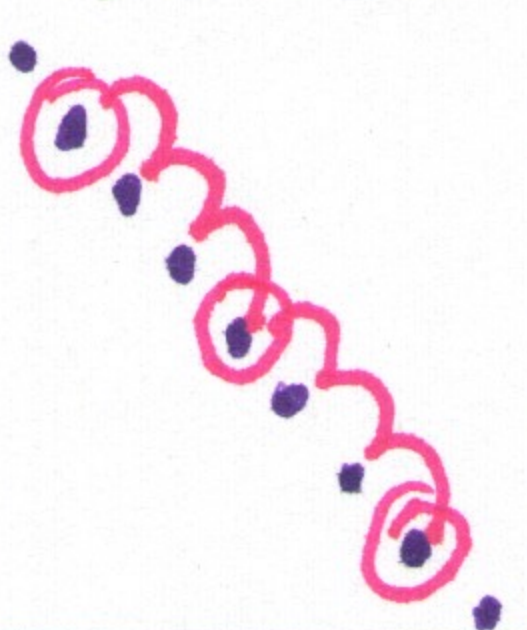f. In which of the above situations can we infer causation? Why?

In b, a randomized, controlled experiment we can conclude causation.

Sampling Methods

We want to survey PCC students on how much they pay for housing per month. Give an example for each type of sampling.

PCC students

Method	Description	Example
Census	ask the whole population	Survey every single PCC student. – student evaluations
Simple Random Sample	Choosing subjects at random – each subject is equally likely to be selected	write a program to randomly select 200 G#'s
Stratified	Divide into large groups and randomly select subjects from each group – may be proportional to the size of the group.	Divide by campus and randomly select a number of students from each campus.
Cluster	Divide the population into many small groups. Randomly select a number of groups. Survey everyone in those groups	Select 10 random CRN#s. Survey all students in those classes.
Systematic	make a list or a line. Choose a random starting point. Select every nth student.	Ask every 5th student in line at the bookstore. Select every 50th G#.
Multistage	combining any of the above methods.	Stratify by campus and select a number of CRN's from each campus. Survey all students in those classes (cluster)

For each scenario, identify the type of sample used.

a. PCC creates an alphabetical list of all currently enrolled students and selects (every 100th student) to survey about their opinion of current tuition rates.

Systematic

b. PCC randomly selects 50 female faculty and 50 male faculty to survey about their thoughts on math education.

Stratified by gender

c. PCC sends an email survey to the [all campus] list serve about food choices on PCC campuses.

voluntary (and convenient)

d. Every student at PCC completes a course evaluation at the end of the term for each course in which they're enrolled.

Census

e. The Portland Water Bureau selects 10 neighborhoods in Portland and surveys (every) member in those neighborhoods about their thoughts on fluoridated water.
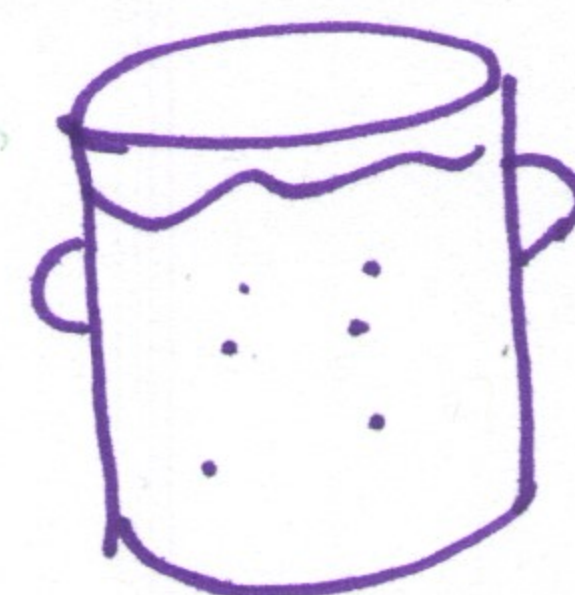
Cluster

f. The first 10 people to arrive on campus on a given day are surveyed about their sleep habits.

Convenience

g. Using G-numbers as identifiers, PCC selects 1000 students at random to complete a survey about their experience at PCC.

Simple random sample

representative sample - doesn't have to be huge

Sampling Methods – Biased Methods

Bias – Any systematic failure of a sampling method to represent the population. A sample is biased if it does not represent the population. There is no way to fix biased data so it is better to design a good survey to begin with.

Methods that are Usually Biased

	Description	Example
Voluntary or Self-Selected Sampling	*Subjects choose whether to participate*	*online polls — only get people from that website* *online reviews — tend to get extreme views*
Convenience Sampling	*any method that is easy or convenient for the researcher*	*Sample my class* *my co-workers* *ask my friends*

Types of Bias

	Description	Example
Selection bias *or sampling bias*	when the sample is not representative.	only survey SE, only call landlines, only call cell phones
Non-response bias	After using a randomized method, someone doesn't participate	instructors selected to give a survey don't participate. Student absent so don't get survey.
Response bias	Anything in the design of the survey that influences the response (answer)	- question order, - not anonymous, - wording of questions

loaded question
leading question

Experimental Design

Chia seeds and weight loss. Chia Pets - those terra-cotta figurines that sprout fuzzy green hair - made the chia plant a household name. But chia has gained an entirely new reputation as a diet supplement. In one 2009 study, a team of researchers recruited 76 participants and divided them randomly into two groups: treatment or control. One group was given 25 grams of chia seeds twice a day, and the other was given a placebo. The subjects volunteered to be a part of the study. After 12 weeks, the scientists found no significant difference between the groups in appetite or weight loss.

Vocabulary

a. Describe the **treatment group(s)** (group(s) receiving each treatment or factor)

the group getting the chia seeds

separate out the psychological response

b. Describe the **control group** (group not receiving a treatment)

the group getting the placebo (fake treatment)

Has **blinding** been used? (**single** or **double**) Has a **placebo** been used?

single — participants don't know which group they are in

double blind — researchers don't know who is in which group.

c. Has **random assignment** been used?

yes

d. What is the **response variable** (what was measured, including units)?

weight loss — measure weight in pounds before + after.

appetite — scale of 1-10 ordinal before + after.

e. Can we generalize the conclusion to the population at large (conclude **causation**)?

There was no significant difference between the groups, but it was a small sample. Maybe measure other variables.

Practice 1. Is exercise effective in combating insomnia? Forty volunteers suffering from insomnia agreed to participate in a month-long test. Half were randomly assigned to an exercise program, while the others did not exercise. Those in the exercise group showed improvement in the quality and amount of sleep.

a. Describe the **treatment group(s)**.

20 people assigned to the exercise program

b. Describe the **control group.**

20 people who did not exercise

Has **blinding** been used? (**single** or **double**) Has a **placebo** been used?

The participants can't be blinded – there is no placebo for exercise. The researchers could be blinded but it doesn't say.

c. Has **random assignment** been used?

yes

d. What is the **response variable**?

quality of sleep – ordinal scale
amount of sleep – hours

e. Can we conclude **causation**?

yes, if the improvement was statistically significant. It was a randomized, controlled experiment, but the sample size is small.

Practice 2. Identify the flaws(s) in reasoning in the following scenarios. Explain what the researchers in each study should have done differently if they wanted to make such strong conclusions.

(a) Students at an elementary school are given a questionnaire that they are asked to return after their parents have completed it. One of the questions asked is, "Do you find that your work schedule makes it difficult for you to spend time with your kids after school?" Of the parents who replied, 85% said "no". Based on these results, the school officials conclude that a great majority of the parents have no difficulty spending time with their kids after school.

This is a voluntary response method, so there will be bias. Parents who have less time are less likely to complete the survey. There could be response bias due to the question wording – it could be worded more objectively. Calling a random sample of parents would give a more representative sample.

(b) An orthopedist administers a questionnaire to 30 of his patients who do not have any joint problems and finds that 20 of them regularly go running. He concludes that running decreases the risk of joint problems.

This is an example of selection bias because the doctor chose patients who do not have joint problems. They should randomly select patients first, then ask whether they run and whether they have joint issues.