

population

sample

What your population is and what your sample is depends on what question you are asking.

eg the students in this class are a sample of

- all USSC students
- all students who have taken AMS 5
- all students in CA
- all students in the US
- etc.

the students in the back row are a sample of the students in this class — which is now the population

Is the sample representative of the population?

If not, the sample is biased.

Sample

Fresh.	18	~ 13%
sophmore.	71	~ 50%
Juniors	27	~ 19%
Seniors	25	~ 18%
	<u>141</u>	<u>100%</u>

back row.

11 ← friends sit together + tend to be of same grade level.
1
2
5

Is this a biased sample or a representative sample?

- it depends on the population

population

all USC students

all students who have taken AMS 5

biased - too many sophmores

representative.

parameter
(of population)

- what we're trying to learn
- the unknown quantity of interest

often hard to measure directly

eg % of voters who would vote for Sanders in the primary election

% of USC students who are female

statistic
(from sample),

- a value computed from a sample which can be used to estimate a parameter.

Data Types.

categorical
(qualitative)

vs

numerical.
(quantitative)

↓
values are names/categories

eg gender
species
colour
party affiliation

↓
deer on campus
thickness of leaves
gpa.

categorical



nominal
no ordering
ice cream flavours

ordinal

rank ordered.



differences
between
categories
does not
have meaning.

numerical

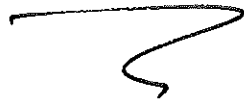


discrete
(quantized).

continuous.

eg
grade points
home runs
deer on campus

eg heights
weights



Controlled Experiments + Observational Studies

controlled experiments → allow us to show causal effects.

eg smoking causes cancer
vaccines prevent disease

observational studies → show association.

may be confounding factors present.

confounding factors - does the vaccine prevent the disease or does better hygiene prevent the disease?

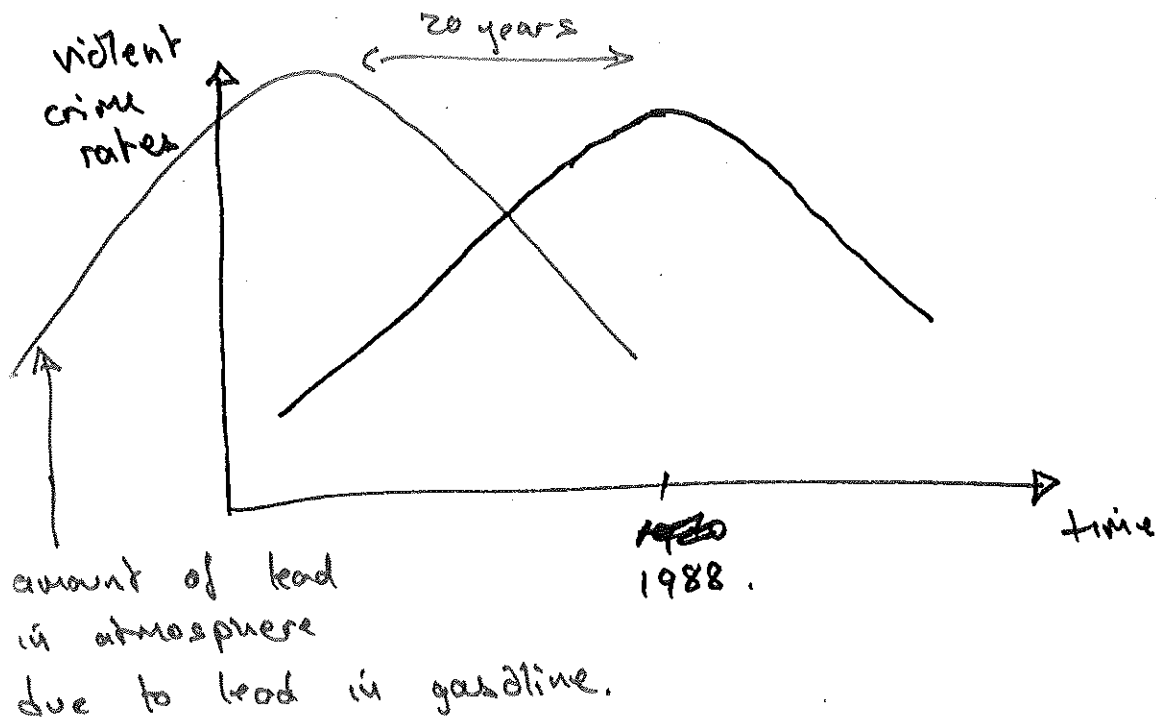
- is there something that we're not measuring that is having an effect?

- do kids from more affluent families score better on tests because they are more intelligent, or because they are better prepared?

Diversion.

Natural Experiment.

- where what should be an observational study can be re-interpreted as an experiment.



Different states removed lead from gasoline at different times.

Different countries removed lead from gasoline at different times.

This lends more weight to lead as a causal factor.

Controlled Experiments.

How to design a controlled experiment

or how to recognize a well-designed one.

Randomized Controlled Double-Blind Trial,

- with explicit hypothesis and a theory for the effect being observed.

Is the treatment effective?

- compare patients who received the treatment with those who did not.

divide the eligible patients into a treatment group, and a control group.

in such a way, that the two groups are as similar as possible in terms of the confounding factors.

- divide Randomly.

- removes bias on part of the person doing the assignment.

- treatment + control - comparisons between the two groups should not be subject to confounding
- the only difference should be the treatment.

- perform the experiment Double Blind

- the patients do not know whether they are in the treatment or control groups ← avoid psychological effects on patient.
- the doctor administering the drug does not know whether they are administering a drug or a placebo avoid giving clues to the patient as to which group they are in
- the doctor assessing the outcomes does not know if the patient was in the treatment or control groups. ← avoid biases in assessment if assessor has an interest in the outcome of the trial.

Case Study : Salk Vaccine Trial.

- eligible patients - some parents refused to allow their children to participate in the study
- these children must be ignored
 - they cannot be added to the control group
 - introduces confounding
reason for refusal may be associated with different disease incidence rates.

randomized controlled - allocate eligible children to treatment / control randomly.

- double blind - all kids got a shot
- either vaccine or saline
 - those performing the assessment didn't know which group the kids were in.

Results.

	size	rate. (cases per 100,000)	
Treatment	200,000	28	
control	200,000	71	
no-consent	350,000	46	← people here are different in important ways from those in treatment/control groups.

Different polio vaccination study.

- give vaccine to all consenting 2nd grade kids.
- used 1st + 3rd grades as control.
- confounds social class (consent/no consent)
- polio is contagious - spreads differently in different grades.

	size	rate
Grade 2	225,000	25
Grades 1, 3	725,000	54
Grade 2 no consent	125,000	44

← lower incidence than in other study, due to confounding factors.