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 ucsb students
    - all students who have taken AMS 6
    - 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:

<table>
<thead>
<tr>
<th>Year</th>
<th>Students</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh.</td>
<td>18</td>
<td>13%</td>
</tr>
<tr>
<td>Sophomore</td>
<td>41</td>
<td>50%</td>
</tr>
<tr>
<td>Junior</td>
<td>27</td>
<td>19%</td>
</tr>
<tr>
<td>Senior</td>
<td>25</td>
<td>18%</td>
</tr>
<tr>
<td>Total</td>
<td>141</td>
<td>100%</td>
</tr>
</tbody>
</table>

Is this a biased sample or a representative sample?

- It depends on the population.

Population:
- All UCSC students: biased - too many sophomores
- All students who have taken AMS 5: representative.
parameter (of population) - what we're trying to learn - the unknown quantity of interest
	often hard to measure directly
	e.g. % of voters who would vote for Sanders in the primary election

% of UCSC students who are female

statistic - a value computed from a sample (from sample), which can be used to estimate a parameter.

---

Data Types.

categorical vs numerical.

(qualitative) (quantitative)

\[ \downarrow \]

values are names/categories
e.g. gender, species, colour, party affiliation

\[ \uparrow \]

# deer on campus
thickness of leaves

GPA
categorical $\rightarrow$ nominal
  no ordering
  ice cream flavours

numerical $\rightarrow$ discrete (quantized)

eg
  # grade points
  # home runs
  # deer on campus

ordinal
  rank ordered;
  differences between categories do not have meaning.

eg
  heights, weights
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.

![Graph showing violent crime rates over time with a peak in 1988 and a decrease in amount of lead in atmosphere due to lead in gasoline.]

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
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.

- remove bias as 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
  - the doctor administering the drug does not know whether they are administering a drug or a placebo
  - the doctor assessing the outcomes does not know if the patient was in the treatment or control groups.
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.

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

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Size</th>
<th>Rate (cases per 100,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>200,000</td>
<td>71</td>
</tr>
<tr>
<td>No-consent</td>
<td>350,000</td>
<td>416</td>
</tr>
</tbody>
</table>

People here are different in important ways from those in treatment/control groups.

Different polio vaccination study:

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

| Grade 2          | 225,000 | 25          |
| Grade 1, 3       | 725,000 | 54          |
| Grade 2 no consent | 125,000 | 44          |

**Note:** lower incidence than in other study, due to confounding factors.