Population

Sample

Parameter - what we're trying to learn
- unknown quantity of interest
- often hard to observe directly
  eg. % of voters who would vote
  for Hillary B.C. in a primary election

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

Data types:

  categorical (qualitative) vs numerical (quantitative)

  values are names/labels
categorical - nominal or ordinal
 ordered
 unordered.

numerical - discrete vs continuous
 (quantized)
 (not).

# deer on campus
# cells

heights
weights
lengths.

---

Controlled Experiments + Observational Studies

Steps of a statistical analysis:

- design the data collection process
- prepare the data for analysis
- describe the data graphically and numerically
- build statistical models for inference.
  - probabilistic model of how you think the data is generated.
- test the validity of the model
- report conclusions
controlled experiments → slow causation
observational studies → confounding factors.

A → B

smoking → cancer

cancer

A → B

genetics

A → B

smoking → cancer

does the vaccine prevent the disease?
does hygiene prevent the disease?

A well-designed experiment can overcome confounding factors.
An observational study (usually) cannot.
Natural experiment:

- Different US states removed lead from gasoline at different times
- Different countries removed lead from petrol at different times

→ Saw same result everywhere

→ We see the same effect for very different populations
In a controlled experiment, we try to set up the groups in a way that all the potential confounding factors are the same (same distribution) in each group.

How to design a good experiment:
(or how to recognize one).

Randomized Controlled Double-Blind Experiment

(A/B test).

Is this vaccine effective?

- compare those who received the vaccine with those who didn’t.

- divide the eligible patients randomly into a treatment group and a control group.
- random assignment eliminates bias on the part of the person making the assignment.

- bias may not be explicit; the person may not be aware they are doing it.

- putting sicker patients into the control group.

**treatment vs. control groups**

- allows comparison between the two groups that differ only by the intervention.

  - comparisons are not subject to confounding.

perform the experiment **double blind**.

- patients do not know whether they are in the treatment or control group.

- person administering the treatment does not know whether they are giving a drug or a placebo.
- the doctor assessing the outcome does not know whether the patient was in the treatment or control group.

1. placebo effect

2. bias by an experimenter with an interest in the outcome.

Case Study: Salk Vaccine Trial

Polio vaccine - when it works in the lab, then test it in larger population eligible patients - some parents refused to allow their children to participate - these children must be ignored as part of the study - they cannot be used as part of the control group, as they may differ in important ways from the participant group.

(polio - class vs. a confounding factor)
randomized controlled - the eligible children were allotted to treatment (control randomly).

(make the groups as similar as possible, except for the intervention)

double blind - all kids get a shot
- some either vaccine or placebo
- doctors performing diagnosis
did not know who was vaccinated.

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
<th>Rate</th>
<th># cases per 100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>200,000</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>200,000</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>No Consent</td>
<td>350,000</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>

- There is a confounding factor associated with giving consent and getting polio.

- not giving consent → get polio

- not giving consent

- get polio

?
Study on 1st/2nd/3rd graders

<table>
<thead>
<tr>
<th>Grade 2</th>
<th>size</th>
<th>rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>225,000</td>
<td>25</td>
</tr>
<tr>
<td>Grades 1,3</td>
<td>725,000</td>
<td>54</td>
</tr>
</tbody>
</table>

Grade 2
no consent

Reduction in rate is smaller.
Confounding has produced a bias
against the vaccine.

How things may go wrong.

- When people do not follow instructions.

Case study: Clofibrate trial

55,552 in one of 5 treatments
2789 to control

Drug reduces cholesterol and should help prevent heart attacks.

Death rate in treatment group 20%
control 21%

- Subjects were not actually taking the drug.
<table>
<thead>
<tr>
<th></th>
<th>Clofibrate</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deaths</td>
<td>Deaths</td>
</tr>
<tr>
<td>Taking</td>
<td>708</td>
<td>1813</td>
</tr>
<tr>
<td>Number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate</td>
<td>15%</td>
<td>15%</td>
</tr>
<tr>
<td>Not taking</td>
<td>375</td>
<td>882</td>
</tr>
<tr>
<td>Number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate</td>
<td>25%</td>
<td>28%</td>
</tr>
<tr>
<td>Total</td>
<td>1073</td>
<td>2795</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>21%</td>
</tr>
</tbody>
</table>

Comparing subjects that took the medicine with those that did not in an observational study:
- Subjects self-assign to treatment/control.

See the same drop in rate in both the treatment and control group between not taking and taking. — Confounding effect.

Conclusions:
- Clofibrate does not have an effect.
- Subjects that take their medication are different from those who don't.

(in ways that affect death from heart attack)
Observational Study.

- When you just watch what happens
- Subjects assign themselves to different groups.

Sometimes it's the best you can do.

Eg accidents
Smoking - unethical to run an experiment.