How Things May Go Wrong.

Random assignment
- make the two groups as similar as possible with respect to the factors that you are not interested in.

If the groups are small, then the randomization may not eliminate the effects of the other factors.

when participants do not follow the instructions.

clofibrate Trial:

5 552 assigned to one of five treatments
2 789 assigned to control

drug reduces cholesterol.

depth rate 20% control group 21%.

- subjects were not actually taking the drug.
<table>
<thead>
<tr>
<th></th>
<th>Clofibrate deaths</th>
<th>Placebo deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Rate (%)</td>
</tr>
<tr>
<td>Taking</td>
<td>708</td>
<td>15%</td>
</tr>
<tr>
<td>Not taking</td>
<td>357</td>
<td>25%</td>
</tr>
<tr>
<td>Total</td>
<td>1105</td>
<td>26%</td>
</tr>
</tbody>
</table>

Comparing subjects that took the medicine with those that did not is an observational study.

- Subjects self-assigned to treatment vs control.

Death rates are reduced by the same amount for those taking the medication, whether it be clofibrate or placebo.

There is something else that differentiates those who take medication vs those who don't, that is having an effect on death rates.
Observational Studies

where you just watch what happens
-subjects assign themselves to the different groups:
  - eg accidents
  - smoking - ethical issues in running a controlled experiment

Observational studies show association
between the factor and response
"linked with"

There may be hidden factors

- genetics

  susceptible
  to cause
  more likely
to smoke

bleach \(\rightarrow\) infectious

-smoker in household

-clean with bleach

-infectious
"Association is not causation"

control for confounding factors
look at subgroups, split up on basis of
hypothesized confounding factors
eg age
sex
smoking

to try to get groups that are as similar as possible in terms of the
confounding factors.

Simpson's Paradox

Relationships between percentages in
subgroups can be reversed when the
subgroups are combined.

Case study: sex bias in UC Berkeley graduate admissions.
8,442 men applied, 44% accepted.
4,321 women applied, 35% accepted.

Appears to be a bias against women.

Is this bias the same across fields of study?

Is the effect still there in subgroups that may be more homogeneous?

<table>
<thead>
<tr>
<th>Major</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Applicants</td>
<td>% admitted</td>
</tr>
<tr>
<td>A</td>
<td>825</td>
<td>62%</td>
</tr>
<tr>
<td>B</td>
<td>560</td>
<td>63%</td>
</tr>
<tr>
<td>C</td>
<td>325</td>
<td>37%</td>
</tr>
<tr>
<td>D</td>
<td>417</td>
<td>33%</td>
</tr>
<tr>
<td>E</td>
<td>191</td>
<td>28%</td>
</tr>
<tr>
<td>F</td>
<td>373</td>
<td>6%</td>
</tr>
<tr>
<td>Total</td>
<td>4,442</td>
<td></td>
</tr>
</tbody>
</table>
Some majors are hard to get into (E, F) some easier (A, B).

Women overwhelmingly applied to the majors that are hard to get into.

*Confounding* of sex + major in the aggregate data.

<table>
<thead>
<tr>
<th></th>
<th>99% A</th>
<th>1% B</th>
</tr>
</thead>
<tbody>
<tr>
<td>51% B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49% A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

80% - party A
20% - party B
Sample Surveys

population - everything that you're interested in
- often too large to record information about every member

use a sample.

1. How do we design the survey that generates the sample to get a representative sample?

2. How to use the data from the sample to make inference about the parameters of the population.

Representative samples eliminate bias.

- selection bias
- non response bias
Selection Bias.

When the people you select for your sample are not representative of the whole population.

- can't detect selection bias just by looking at the data in the sample
- important to construct the sample in a way that is unbiased.

Example: 1936 Election

Loyalty Digest held a poll

<table>
<thead>
<tr>
<th>Prediction</th>
<th>Roosevelt</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>43%</td>
<td>62%</td>
</tr>
<tr>
<td></td>
<td>57%</td>
<td>38%</td>
</tr>
</tbody>
</table>

↑ from a biased sample.

How to choose a good sample:

- randomly from the entire population

What did they do?

surveyed 10 million people by mail

addresses came from telephone directories, club mailing lists.
‘1 in 2’ admits to plagiarism

Michael Stratford
Chief News Editor

Half of Cambridge students have committed plagiarism, as defined by the university according to a Varsity survey. 49 per cent of students admitted that they had plagiarised work, although this differed radically between subjects and colleges. Biologically trained students of the Law Faculty plagiarised the most out of any subject, with 62 per cent of them breaking the university rules. The second highest was the Archaeology and Anthropology department with 59 per cent.

"It is a depressing set of statistics," said Robert Doley, a tutor in biological Anthropology at King's College.

The college at the bottom of the Tompkins' table, St Edmund's, had the highest proportion of plagiarising students, with 67 per cent admitting to breaking the university rules. Scholastic at the top of the Tompkins' table had the lowest number of plagiarising students.

"It stands to reason that those students are performing less well with more work underhand means to get by," said a member of the University Council, the principal executive and policy-making body of the university.

It is perhaps not surprising that 80 per cent of students said that the university is doing enough to punish plagiarism. "You can see why students do not need to break the rules to their own benefit would be keen to uphold the impression that the system is working," said a member of the General Board, the body responsible for policy at the University.

"Sometimes when I am really fed up," said a Law Economy student at Pembroke, "I Google the essay title, copy and paste everything into a blank Word document, and add another bit. They usually end up being the best essays," 100 per cent of students admitted to copying without acknowledgement. But the results should be taken lightly because less than five per cent of the students who admitted to plagiarism.

Continued on Page 3

Cambridge application levels hit record high of 15,000

Vicky Winson
News Reporter

The University has announced that the number of applicants for undergraduate courses has reached a record high this year. For the past four years, around 13,500 students have applied to study at Cambridge, whereas this year the figure is set to be closer to 15,000. This equates to an increase of around 12 per cent, compared with an average of only 6.5 per cent across all UK universities.

Some students have faced better news than others. Computer Science has seen a 53 per cent rise in applications, with the next highest rise being in Philosophy which is up by 43 per cent. Theology and A.S.NaS, saw applications drop by 18.8 and 16 per cent respectively.

This rise in applicant numbers comes at the end of a year in which many changes have been made to the admissions process. The requirement for a foreign language GCSE has been removed because of fears it disadvantages applicants from state schools, where foreign language teaching has declined rapidly in recent years. The Cambridge Application form was also abolished, the form which students had to fill in alongside their UCAS application, was felt to be off-putting for some students. Continued on Page 3

49% Percentage of survey respondents who admitted to some form of plagiarism.
62% Percentage of respondents studying Law who admitted to plagiarism.

Continued on page 3

Richard Eyre, former Director of the National Theatre, on the ADC Goldfrapp reviewed

Continued on page 3
The Original Classic with a
New Introduction by the Author

More than
48 million
copies sold
worldwide

THE
HITE
REPORT
SHERE HITE

A Nationwide Study of
Female Sexuality

More than three thousand women describe to each other what they
and other feelings about sex, including:
- What they like and what they don't like
- How urgent is the need—sex with or without menchemistry
- How it feels to be free, and not free during and after sex
- The importance of sexual satisfaction and masturbation
- The greatest pleasures and frustrations of their sexual lives
In 1936, telephone ownership was concentrated amongst the wealthy.

1936 was during the great depression.

→ poor people vote differently than wealthy people, especially during a depression.

Taking a large sample with a biased procedure does not improve the results.

**Non-Response Bias.**

*Hike Report:*

98% of respondents were unhappy with their marriage.

72% had extra-marital affairs.

Only 4% of surveys were returned.

- people who are unhappy with their marriage are more likely to reply.

Later random sample: 93% satisfied.

7% having affairs.
How to choose a good sample:

- use simple random sampling.
  - decide how large a sample is needed
  - from the list of the entire population
    - draw an element at random
    - delete that element from the list
    - repeat until you have the desired number of elements in the sample

but: often hard to do
  : often pretend we're doing this, as it makes the math easier.

other types of sample:

Quota Sampling / Stratified sampling
  - try to match the sample to the population
eg: specificity
  # women + # men
  age distribution of men / women
  race of men + women
  etc.
need to be careful not to introduce biases
- eg if interview people in downtown SC
  at 11am on a Tuesday, the women will
  likely include more women who do not work outside the home than
  is representative of the general population.

Cluster Sampling.

Generate a random sample, where the people in the sample are not as
geofraphically spread out as those in a
simple random sample.
Finally, the actual person to be interviewed is determined in advance.

- e.g., youngest woman over the age of 18.

- removes bias - selection bias or part of interviewee
  - selection bias in terms of who is home.

Allows for smaller errors with smaller samples.

Telephone Surveys:

generate phone numbers randomly.

- non-response bias.

- selection bias.
  - not include cellphones.
  - biases the sample towards older people (who have land lines).