

## 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

5552 assigned to one of five treatments  
2789 assigned to control

drug reduces cholesterol.

death rate 20%  
control group 21%.

- subjects were not actually taking the drug.

	clofibrate deaths		Placebo deaths.	
	number	rate (%)	number	rate
taking	708	15%	1813	15%
not taking	357	25%	882	28%
total	1103	20%	2789	21%

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

- subjects self-assign to treatment / 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

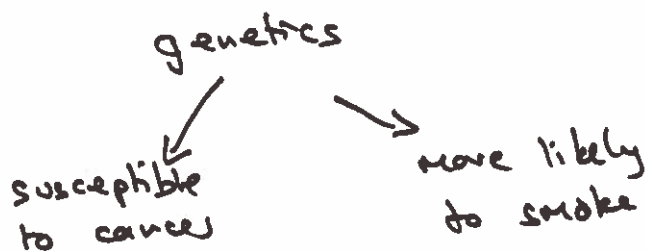
smoking

- ethical issues in running  
a controlled experiment

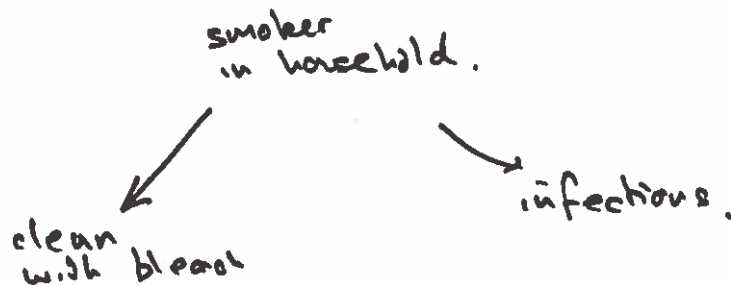
observational studies show association  
between the factor and response

"linked with"

There may be hidden factors



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.

8442 men applied, 44% accepted.

4321 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?

major.	men		women	
	# applicants	% admitted	# applicants	% admitted
A.	825	62	108	82
B	560	63	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	373	6	341	7.
total		44		30.

↑  
women admitted at higher % for most majors.

Some majors are hard to get in to (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.

---

	99% A	
51% B	1% B	
49% A		

80% - party A  
20% - party B

---

## Sample Surveys

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

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 ~~see~~ 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

Literary Digest held a poll

prediction	Roosevelt	43%	actual	62%
	Landon	57%		38%

↑  
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.



# VARSlTY

Friday October 31 2008

The Independent Cambridge Student Newspaper since 1947

Issue No 681 | varsity.co.uk

»p14 Arts  
**Halloween-themed comic strip**

»p12 Features  
**American Election Special Preview**

»p34 Sport  
**Legendary writer Simon Barnes**



## '1 in 2' admits to plagiarism



» Varsity survey reveals that 49% per cent of Cambridge students have committed some form of plagiaristic act whilst at the University  
» Students criticise unrealistic plagiarism guidelines; others blame rigour of Cambridge degrees

Michael Stothard  
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 have plagiarised work, although this differed radically between subjects and colleges. Ironically, 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 Foley, a Professor 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. Selwyn, at the top of the Tompkins table, had the fewest number of plagiarising students.

"It stands to reason that those students who are performing less well will resort to more 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, a great number of whom are frequently breaking 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 education policy at the University.

"Sometimes when I am really fed up," said Land Economy student at Pembroke, "I Google the essay title, copy and throw everything on to a blank word document and jiggle the order a bit. They usually end up being the best essays." 100 per cent of Land Economy students admitted to plagiarising, but the results should be taken lightly because less than five per cent of the student population replied to the survey. *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

## Cambridge application levels hit record high of 15,000

Vicky Woolley  
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 rise of only 6.5 per cent across all UK universities.

Some subjects have fared better than others. Computer Science has seen a 53 per cent rise in applications, with the next biggest rise being in Philosophy which is up by 43 per cent. Theology and ASNaC saw applications drop by 18.5 and 10 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 disadvantaged 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*

»p16 Richard Eyre, former Director of the National Theatre, on the ADC »p22 Goldfrapp reviewed

*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 their own words their  
most intimate feelings about sex, including:

- What they like and don't like
- How orgasmic they like—why and without intercourse
- How it feels to be horny and when it's going on
- The importance of sexual stimulation 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.

### Hite Report:

98% of respondents were unhappy with their marriage

72% ~~were~~ 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 specify

# women + # men

age distribution of men / women

race of men + women

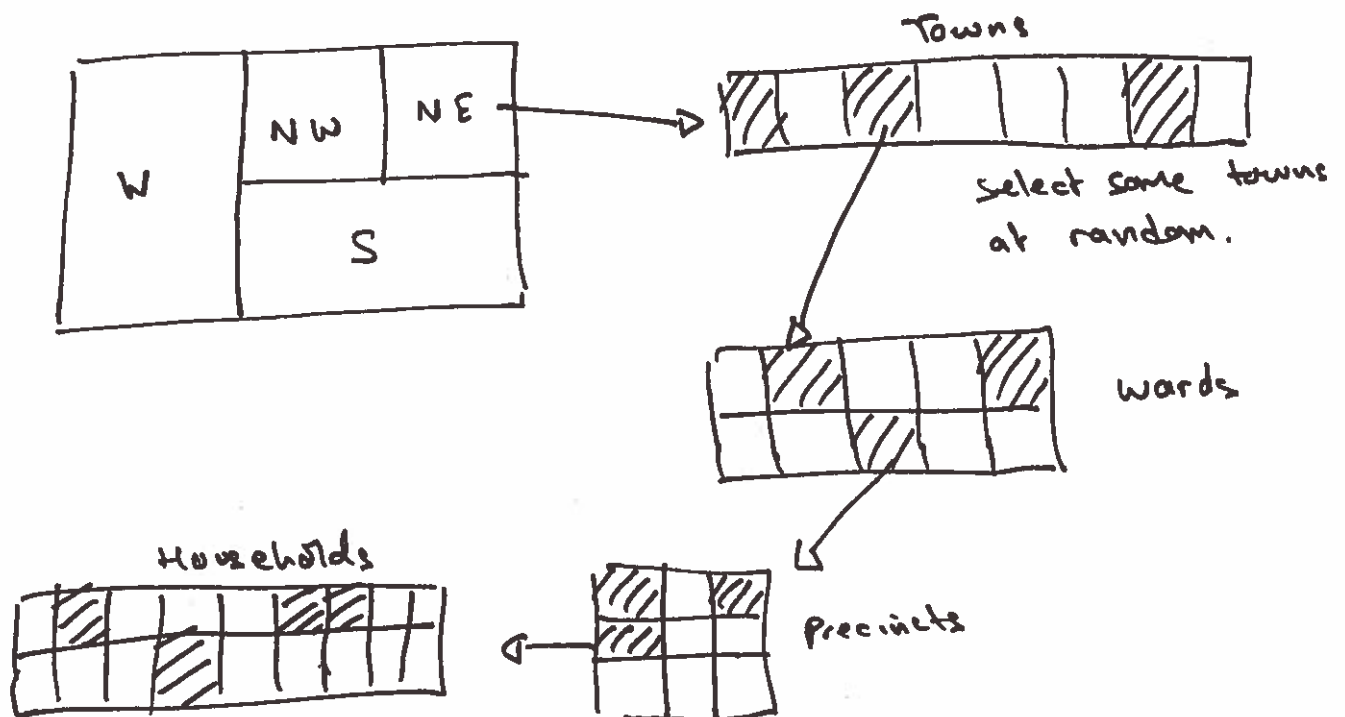
etc.

- 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 ~~more~~ women who do not work outside the home than it is representative of the general population.

### Cluster Sampling.

Generate a random sample, where the people in the sample are not as geographically spread out as those in a simple random sample.



Finally, the actual person to be interviewed is determined in advance

- eg, youngest woman over the age of 18.

- removes bias
- selection bias on part of interviewer
- 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).