

§3.1 Design of experiments.

observational : here is some data we found! let's look at it...

- easy to find correlations, hard to know causation.

example : dramatics
correlated w/ ice cream sales.

experiment : we have some idea we want to check.

↑ confounding
 lurking variables.

- e.g. take 20 patients, give 10 the drug see if it makes any difference.

↑
subjects.

10 placebos

↑ choose at random!
treatment.

more response.
eg. - live die

random assignment : label 1, 2, ... 20

choose 10 at random, eg pick random number between [0,1] in R `runif(1)`

scale up chosen that, delete repeat w/ 19.

Warning sometimes this is hard! Choose 10 votes in Staten Island at random?

Experiment : want . control. (usually by comparing two treatments).

- randomize
- have enough subjects.

If the observed effect is large enough that it is unlikely to occur by chance, then we say it is statistically significant.

Warning still hard to treat people the same: double blind neither patients nor doctors know who gets placebo.

§3.2 Sampling design

population : everyone we care about, e.g. voters in US.
thing

sample : some smaller subset we collect data about.

choosing random samples is hard : many samples are biased :

convenience : I ask my friends.

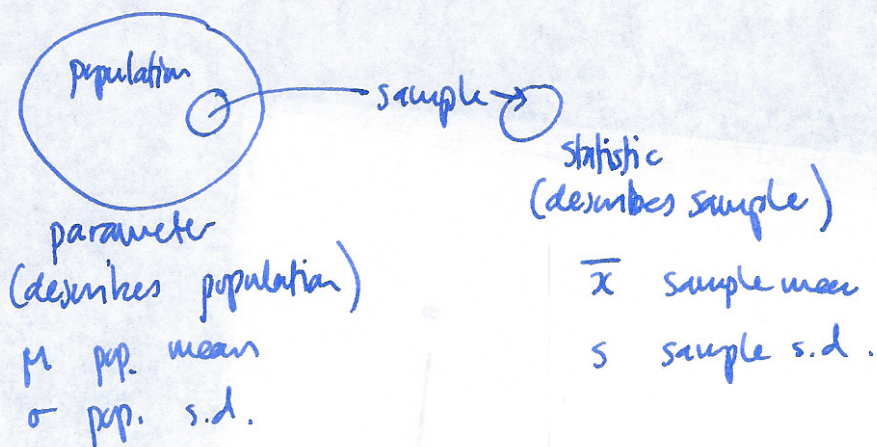
voluntary response : people who are biased.

simple random sample of size n : choose n people at random from population.

stratified sample: people: $\rightarrow$ men $\leftarrow$ sample from each one.
 women $\leftarrow$

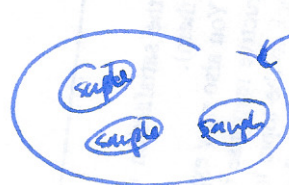
⑥

§3.3 Statistical inference.



Q: what does the sample tell us about the population?

sample variability:



sample distribution: distribution of sample statistics.

Bias vs variability:



reduce bias: use random sample

reduce variability: use larger sample.

[variability of sample statistics depends on sample size, but not size of pop as long as $n > 100$ (bigger)]

§3.4 Ethics

studies: individuals give informed consent } IRB.
 ind. data kept confidential