

# Sampling

Wenda Zhou

June 1, 2017

# Sampling

- ▶ First step in statistical analysis: collecting data
- ▶ Interested in how one variable (dependent) varies according to some treatment
- ▶ Two main paradigms: observational studies and experiments

# Experiment

- ▶ Researcher controls treatment assignment
- ▶ Provides most evidence

## Randomized trial

Usually have treatment assignment random.

Patients under treatment are “statistically same” to patients under control

# Experiment

## Control group and placebo

- ▶ Inference only possible through comparison, so have treatment and control group.
- ▶ Wish to minimize difference between treatment and control group – placebo

# Experiment

## Sub-populations

Experiments give strong evidence for experimental population.  
Need to ensure that experimental population corresponds to larger population.

# Observational studies

Not always possible to have experiment:

- ▶ Impossible (e.g. effect of gender on salary)
- ▶ Ethical difficulties (e.g. effect of diet on health)
- ▶ Data already collected

# Observational studies

Treatment assignment is assigned through uncontrolled mechanism

- ▶ Prospective: participants selected a priori then observed.  
Better evidence.
- ▶ Retrospective: participants selected post-hoc.

# Prospective study

## Framingham heart study

Ongoing study since 1948. Tracking about 5000 adults and their descendants.



# Retrospective study

## Example

Collect medical history of 100 smokers and 100 non-smokers.

Numerous advantages over prospective studies:

- ▶ Cheaper and faster – data already exists
- ▶ Advantages for rare events (e.g. lung cancer)

# Confounding

Observational studies can face large problem of confounding.

## Diamond prices

Price and color confounded by weight

## Sunscreen and skin cancer

Sunscreen usage and skin cancer both driven by exposure to sun.

# Natural experiments

Observational studies are easier but provide weaker evidence.  
Sometimes can have “natural experiments” in observational studies.

# Natural experiments

## Examples

### Twins separated at birth

Separate genetic vs environmental effects.

### Vietnam draft lottery

Separate effect of volunteers