Lecture 14: Sampling

Dewayne E Perry
ENS 623
perry@ece.utexas.edu

Sampling Information

- Must describe sample sufficiently
  - To judge representativeness
  - To evaluate equivalence of different groups of participants
  - To assess whether participant variables have been controlled
  - Enough details to compare with other studies
- Representativeness
  - In many cases simply assumed
    - Eg, the populations fir the major categories
    - Assumed sample in Chicago behaves the same as sample in NYC or London
  - In surveys, representativeness critical
    - Eg, Roosevelt/Landon election
      - ✓Predicted for Landon
      - ✓Huge sample for prediction
      - ✓BUT from car/telephone owners
      - ✓Biased towards conservative and hence Landon

Non-probability Sampling

- Self selected sampling
  - Eg, in media: open invitations to respond to questions
  - Sampling limited to those who saw the request
    - Eg, people with computers
  - Unclear whom any the self-selected surveys represent
  - Slightly different: consumer's reports
    - Select population
    - Self-selected within that
- Haphazard sample
  - Recruiting in public space – eg, airports, malls
  - Difficult to replicate
  - Danger of biased samples
    - People who travel airlines don't go to laundromats, etc
    - But some topics don't make a difference: optical illusions
    - For social attitudes, bias could be pivotal
  - *NOT* random sampling

Non-probability Sampling

- **Sample of opportunity**
  - Convenience samples
    - Pool of participants who are available
      - ✓ Eg, this class
    - For the psychologist might be ok
  - The more the dependent variables are associated with variables other than the independent variables the more crucial representativeness becomes
  - A variation: participants from a hospital, school, clinic, project etc
    - May have more than enough people to meet requirements
    - And random selection from population
    - May need to know how representative population is

Non-probability Sampling

- Homogeneous, restricted, purposive sample
  - A specific subset of a convenience group
    - Can restrict homogeneity wrt feature shared
    - Eg, personal adjustment patterns in freshmen
    - But expected to represent all freshmen, not just Princeton or exclusively Caucasians
  - Eg, bias effects if too restrictive a sample
    - SAT national average: 906; Mississippi: 1001; NJ: 889
    - Only 3% take it in Mississippi; 65% in NJ
- Networking (snowballing) sample
  - Ask for references if do not have enough samples
  - Eg, networks of mothers who have small children
  - Can suffer from inbreeding, ie, too homogeneous
  - Vulnerable to contamination of the results if participants talk with each other about the experiment
- Systematic sample
  - Eg, first 50 people thru the door
  - But early arrivals may differ from latecomers in systematic ways

Probability Sampling

- Random sampling
  - Each member has an equal chance of being selected
  - large population often beyond scope of most researchers
  - From more limited populations is possible, but have to be careful
- Systematic sampling
  - Often hoped that systematic methods are unbiased and equivalent to random
    - Have to be careful that no bias introduced
  - Eg, every third person on list

Probability Sampling

- **Stratified sampling**
  - Usually limited opportunities, unless well-funded
  - Costly to represent/match target populations on all demographic and other variables
  - Physical scientist can assume 1 oz of silver representative worldwide
  - By psychologist studying female depression has to worry about a host of demographic and personal variables
- **Cluster sample**
  - Randomly target clusters of people
    - Eg, students in schools in a city
    - Then randomly within schools

Sample Sources

- Direct samples
  - Obtain data directly from people in sample
  - Experimental, quasi- and non-experimental commonly obtain data this way
- Archival samples
  - Use data already gathered and are a matter of record
    - Eg, actuarial records such as vital statistics, medical records, etc
  - Experimenter bias can have no influence on them
  - Disadvantages: forced to rely on accuracy and timeliness of data
    - Eg, income of 20 years ago not very useful today
    - On test records, have no control over qualifications of examiner
    - Or accuracy of scoring, administration, or interpretation
  - Advantages: not contaminated by the experiment – more than balances
  - Other disadvantages
    - Selective deposit, survival
    - Selective entry factors may distort

Sample Biases

- Volunteers
  - Accept or decline may bias sample
  - May not be representative of population
    - Unrepresentative characteristics
    - Could threaten generalizability
  - Ethical issues
    - Need sufficient info for informed decision
    - Foreknowledge can cause problems
  - Levels of volunteering
    - Anonymous opinion of social issue
    - Participation where no noxious effects
    - Out of their way, extra time, some discomfort
  - Declining may be a function of commitment
  - Recompense could effect level as well
  - Greater the sacrifice, fewer volunteers
  - Cannot know characteristics of non-volunteers

Sample Biases

- Rosenthal & Rosnow 1991
  - Maximum confidence: tend to be better educated, higher in social class, more intelligent, more approval motivated, and more sociable
  - Considerable confidence: see arousal, be unconventional, female, Jewish, non-authoritarian, nonconforming
  - Some confidence:from smaller towns, interested in religion, more altruistic, more self-disclosing, more maladjusted, younger

Sample Biases

- Biasing by selective attrition
  - At beginning sample may be representative
  - Attrition may cause non-representative
    - Dropouts are mainly women, middle aged, poor, etc
  - Obligation to exclude those no longer willing or able causes problems
    - What about people who do not respond to treatment
    - Eg, people who are not stressed by stress condition
  - Have we lost randomness as a result?
  - Rationale for exclusion should be made clear

Assignment

- Random
  - Selection brings into study, assignment places them in treatment
  - Does not solve problem of non-equivalent groups
  - Randomness by random number table simplest
- Systematic
  - Potential for bias always present
  - May have confounding variable present
  - Must convince two groups are equivalent

Assignment

- Sample size
  - As many as possible dwindles quickly to as many as feasible
  - Significant results can be obtained with 20-30 participants, 10 per treatment, provided
    - Distributions are reasonably normal
    - Statistical assumptions are met
  - Inconclusive results: sample size problem?
  - Large sample -> small differences could be significant
  - Power analysis: increase power by
    - Raising level of significance required
    - Reducing standard deviation
    - Increasing magnitude of effect by using strong treatments
    - Increasing the size of the sample
  - $p < .05$ usual desired level of significance