The Logic of Experiments

Soc 357
Fall 2006

Key Components of a "True" Experiment

1. Manipulable independent variable a.k.a. A treatment variable
2. Experimental Control
   1. Control of experiment setting
   2. Control of extraneous variables through random assignment to treatment & comparison groups
3. Measured Dependent Variable

Quasi- and Pre-Experimental Designs

- ***Quasi experiment*** = manipulable independent variable but no randomization
- Quasi-experimental design important in applied research where random assignment of treatments is impossible or unethical
  - Eg. Doob and Gross, "Status of Frustrator"
  - Medical research
- ***Pre-experimental*** designs: lack comparison group, or random assignment to comparison group
  - Eg. Classroom behavior example

WHY CAN'T WE INFER CAUSALITY from these?

1. Manipulable Independent Variable

- The essential key to any experiment or quasi-experiment
- The researcher can control whether a subject gets a particular treatment or not
  - "treatment" = category of independent variable
- ***Manipulable***: things the researcher can do such as give a treatment, confront the subject with some experience
- ***Not manipulable***: characteristics the subject is born with OR attitudes, actions, characteristics under the subject's control

Experimental Control and Causal Relations (1)

- Recall the four criteria that need to be met in order to theorize about causation:
  - **Statistical association**: two things vary together
  - **Direction of influence**: Cause precedes effect in time
  - **Elimination of rival hypotheses:** Other variables are ruled out as possible explanations for the relationship
  - **We can identify the mechanism** for the cause-effect relationship, we know how it works

Experimental Control and Causal Relations (2)

- In "true" experiments, it's easier to rule out all other possible explanatory factors ("elimination of rival hypotheses") by controlling the conditions of the experiment.
- In a perfectly executed experimental design, there is only one alternative explanation for some association we see: that it happened by chance.
- We assess this using statistical ***tests of significance*** - that tell us the probability that the association we're seeing is just by chance.

Experimental control of a setting

- Control of all aspects of a research setting
- Like putting on a play: everything is scripted
- Only one thing changes from subject to subject: the independent variable/treatment
- Everything else in the setting remains the same

Settings: Field & Laboratory Experiments

- Most experiments take place in a laboratory: easier to control setting
- Field experiments meet all the requirements of a "true" experiment, but staging occurs in a natural environment
  - Eg. Piliavin et al. misdirection study
  - This is what you'll be doing!

Comparison Groups

- We use comparison groups to isolate the effect of the variable we are interested in – everything is exactly the same in both groups except our independent variable
- Can organize comparisons in two ways:
  1. Within-subjects
  2. Between-subjects

Comparison groups: Between- and Within-Subject Designs

- ***Within-subjects designs***: Each subject gets all treatments. A subject acts as its own comparison group. Powerful design IF
  - Feasible
  - Treatments are reversible (so you can run ABA designs)
  - No interference between treatments (can minimize effects of time)
- ***Between-subjects designs***. Each subject gets only one treatment. Randomization or other procedures equate the control groups
- Our projects will be between-subjects.

Within Subject Example: Discrimination Studies

- Mary is black. Applies for a job at a shop on State Street. Is not given an application, told they are not hiring. Is this evidence of discrimination?
- What if Jane who is white goes there half an hour later and is given an application?
- In this case **the employer is the subject**, and the treatments are the race of each person applying
- Common in a variety of discrimination studies
- Issues of controlling for other factors, reactivity to treatment; ethics

Between-Subjects Example:

- Rubin, “Measuring Romantic Love”

  - Experiment randomly assigned couples to a “together” group or an “Apart” group to discuss dating & marriage

  - Within groups also had “strong” and “weak” pairs, based on scores on questionnaire.

- TWO independent variables: setting of discussion, strength of relationship, FOUR comparison groups

Factorial Designs

- More than one independent variable
  - Eg. UW Sociology grad student Devah Pager, "The Mark of a Criminal Record"
  - Audit study that tested both race & having a criminal record on being called back for a job interview
  - Sent matched pairs of males to apply for jobs with made-up resumes
    - Pairs matched on race, age, education, work experience, self-presentation
    - Within pairs, differed on criminal record

"The Mark of A Criminal Record": Results

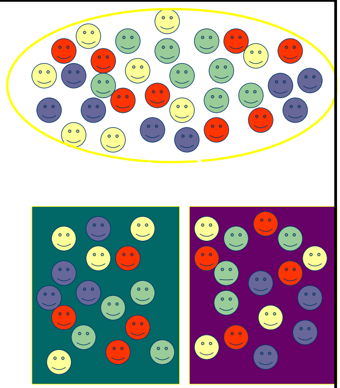
Percent of applications getting employer call-backs

	Criminal Record	No Criminal Record
White	17%	34%
Black	5%	14%

Experimental Control of Extraneous Variables

- You can control setting but you can't control all of your subjects' qualities (mood, temperament, physical wellbeing)
- These could be producing the effect you see, not **your** independent variable
- In between-subjects designs, we get around this by using ***randomization*** and ***holding constant***

Randomization



- Take a group of subjects, and randomly divide them into treatment groups (no cheating!)
- Use a true random device, eliminate any person's choice about who is in which group

The "Magic" of Randomization

- All subject characteristics are statistically equated automatically, whether or not you can list them
- Selection bias automatically controlled
- Depending on design, other factors may be controlled too.

Randomization equates treatment groups

- This works ***in the statistical long run***
- Depends upon sample size
- EXAMPLE: Assume you have two "types" of people (I.e. males, females) that are distributed 50-50 in your subject pool
  - What are the chances that randomization will produce a pretty good division, I.e. that the male-female mix is no worse than 40-60 in each group?

Sample size and effect of randomization

- The next few slides show how sample size affects this
- Sample size is the total number to be divided into treatment groups
- The charts show the distribution of proportions “male” among all possible random divisions of a sample into two groups

Sample size 4, 2 cases per treatment

Population: Ma Mb Fa Fb

Six possible ways they can be divided randomly into two groups of two:

IV= 1	IV= 2	% split
Ma Mb	Fa Fb	100-0
Ma Fa	Mb Fb	50-50
Ma Fb	Mb Fa	50-50
Mb Fa	Ma Fb	50-50
Mb Fb	Ma Fa	50-50
Fa Fb	Ma Mb	0-100

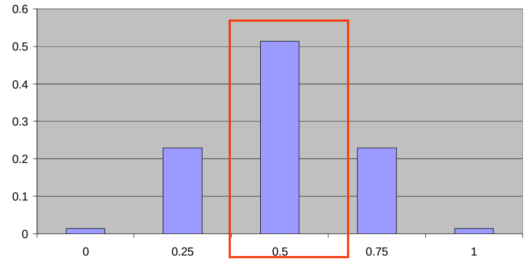
Split	Probability
100-0	1/6 = .17
50-50	4/6 = .67
0-100	1/6 = .17

Sample size 8, 4 cases per treatment

There are 70 distinct ways 8 people can be randomly assigned to two groups of 4 each.

IV 1	IV2	# ways	Probability
MMMM	FFFF	1	1/70 = .014
MMMF	FFFM	16	16/70 = .229
MMFF	MMFF	36	36/70 = .514
FFFM	MMMF	16	16/70 = .229
FFFF	MMMM	1	1/70 = .014

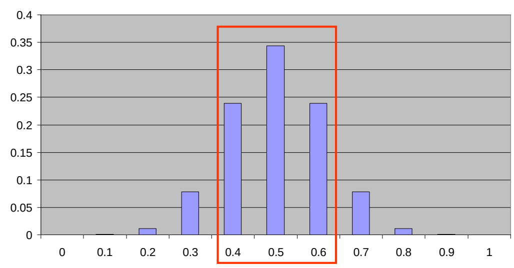
Sample size 8, 4 per group



Probability is .51 of an even M/F split in the randomization.
Probability is .028 of all M in one group and all F in the other

Sample size 20, 10 per group

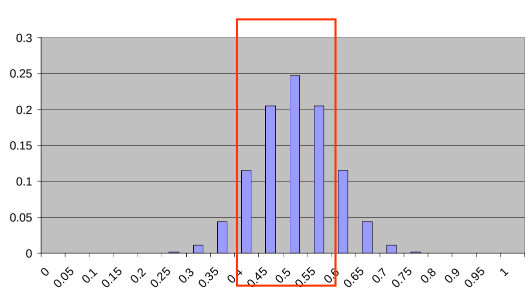
There are 184,756 possible random divisions



Probability is .821 of 60-40 or better M/F split in the randomization.
Probability is .023 of a 80-20 or worse split

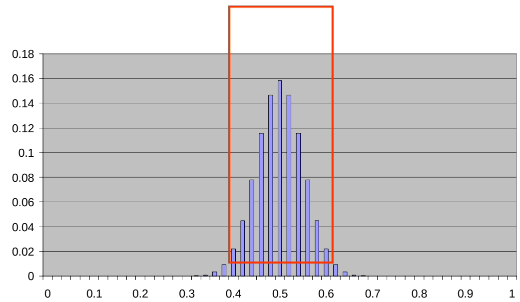
Sample size 40, 20 per group

There are 137,845,229,220 possible random divisions



Probability is .887 of 60-40 or better M/F split in the randomization.
Probability is .00036 of a 80-20 or worse split

Sample size 100, 50 per group



Probability is .973 of 60-40 or better M/F split in the randomization.
Probability is 0.000000002222 of a 80-20 or worse split

Sample size in our projects

- Minimum is 10 cases per treatment, 20 cases total for 2 treatments (30 cases total for 3 treatments)
- This does not ASSURE a “good” division of subjects, but makes it quite probable (better than 80% for one dichotomy that is distributed 50-50 in the population)
- If your experiment is easy & fast, 15 or 20 cases per treatment is even better

Holding Constant vs. Randomization

- Holding Constant: subjects exactly the same. Absolute control of extraneous variables held constant.
- Randomized: there is variation, but can be expected to even out in the statistical long run
- Constant: no generalizability to other characteristics.
- Randomized: more generalizable

Example: Holding Gender of Subject Constant vs. Randomized

- Gender of subject (subject characteristic you cannot manipulate)
- Constant = men only (or women only). Total control, but cannot generalize to the other gender
- Randomized = both genders, let randomization equal it out. Some risk of uneven balance if sample is small. Can generalize to both genders.

Matching

- Another way of ensuring comparability of two groups
- Identify pairs of people who are similar on some characteristic you think is important
- Randomly assign them to either group
  - Eg. Discrimination audit studies – pairs of people are matched up on all characteristics except race (age, income, education, work experience, etc.)

Randomization: Concluding Points

- As sample size increases, probability that randomization produces about equal groups goes up
- We use matching, holding constant to achieve better distribution of salient characteristics
- But there is still the problem of unidentified characteristics making a difference
- So sample size matters

Randomization ≠ Random Sampling

Randomization	Random Sampling
Internal validity (isolating causal relations)	External validity (generalizing outside sample)
How you divide a pool of subjects into categories of IV	How you get a pool of subjects
Can be done to a convenience sample	Random samples from same population can be equivalent of randomization

Uncontrolled Extraneous Variables

- What's left?
- Other differences between categories of IV besides what you have in mind
  - How long a treatment takes
  - Personality characteristics of person embodying experimenter characteristics (e.g. sex, race)
- Distortion in measure of DV that varies by IV. Possible if DV is subjectively coded and person coding knows the IV
- Environmental disturbance that affects one category of IV more than the other

Limits of Experimentation

- Cannot experiment on subject characteristics (e.g. race, sex, family background) which people are born with. Physically impossible.
- Socially impossible or unethical to experiment with socialization history of children
- Issues of ***External Validity***: can you generalize to a larger population?

Examples of Experiments

- Social psychology: e.g. Darley & Batson
- Medical treatments: IV=medicine, DV=symptoms
- Social treatments, e.g. reduce fertility in Indian villages: IV= propagandize women only vs. couples together, DV= number of children born)
- Discrimination studies: IV= race or sex of person applying for job or housing, DV=how the person is treated

Doing the Experiment

Examples

- Asking directions (bridge, SERF)
- Correcting misinformation X modeling: elevators, subway
- Putting on some kind of scene, how do people respond? (Darley & Batson an example)
- Phone scenarios: will they make a call? [caller ID may reduce]
- Paper & pencil experiments on perceptions (Goldberg an elaborate example)
- Petitions - # of people already signed is manipulated
- Whether people agree to be in a survey X what it is about

Keys to Experiment Project

1. Manipulated independent variable
2. Measure of dependent variable that is well-defined, is the same for all categories of IV
3. Hold constant everything else in the setting *except your IV*. Especially control your own behavior with a script etc.
4. Randomize subjects and time. Balance out in statistical long run the things you cannot hold constant.

Manipulated Independent Variable

- MANIPULATED independent variable = experimenter controls which category of independent variable subject can fall into
- CANNOT do experiments with subject characteristics as independent variables
- E.G. Sex of subject (not manipulated) VS sex of person making request (which may be manipulated)
- Experimenter decides who is in which treatment group

Examples of manipulated variables

- Sex, race, age etc. of person making request
- Something about what you say or do
  - Different explanation or instructions
  - How you act
  - Whether someone says something
- What you are wearing (e.g. UW shirt vs. Ohio State shirt OR shirt & tie vs. sweatshirt)
- Words in a document (e.g. names of authors or characters, changes in wording)

Dependent Variable

- You do not control the subject's behavior, you observe it
- Can be field observation (e.g. helping the victim)
- Can be a small survey (e.g. ask them a few questions)
- Can be whether they agree to do something
- etc

Examples of DVs

- Helping: picking up dropped books/groceries, answering a question
- Making a phone call or not
- Agreeing to be in a survey or to sign a petition or to do something
- Answers to specific survey questions
- How long it takes to do something

Measure of the DV must be consistent

- You MUST use exactly the same measure of the dependent variable for ALL treatments (categories of independent variable).
- All possible categories/responses of the DV must be logically possible for all categories of the IV
- You need to devise a way of measuring the DV that can be done reliably and that can fit into the setting.
- Have to specify the operationalization of the DV. May be simple or complex, depending on what it is

Unit of Analysis

- May be person, or may be trial
- Person: you randomly assign treatments to individuals and record individual behavior. The DV is the behavior of a particular person.
- Trial: you randomly assign treatments to units of time and record the behavior of any/all people in the setting. The DV is the behavior of everyone in the setting.
  - Elevator rides: does anyone in the elevator correct misinformation, or complain?

Hold everything else constant (except what you randomize)

- Have a script so you say/do the same thing on every trial (except for DV)
- Plan answers to common questions so they will be consistent
- If a paper experiment, papers are identical except for the IV
- Make as small a change as possible for the IV and have everything else be the same

Plan & practice

- You are responsible for doing prior planning/practice to get everything held constant except your IV
- Pretest is essential: try out your procedures to see if they will work before your main test
- OK to hold some subject characteristics constant (I.e. only men, only people who answered the phone, only people who appear to be students)

Experimental Control

- Hold constant or randomize everything except the IV
- May hold some subject characteristics constant, but must always randomize the others – impossible to hold everything constant

Doing Randomization

- Method 1 (best)
  - FIRST pick subject (any way you want)
  - THEN randomly assign IV to that subject
- Method 2
  - Randomly pick next value of IV
  - Select subject according to impersonal mechanical rule. Must keep all subjects in the data.
- ***Any person who has been "exposed" to IV MUST be kept in the data.***

Doing Randomization Right

- Real randomization: little pieces of paper pre-counted
- Key is how this relates to the subject selection
- If you randomly choose treatment but then non-randomly choose which subject gets the treatment, you have destroyed the randomization
- Need to explain how the randomization fits into the whole flow of the experiment.

Mixture of Constant and Random

- TIME: “Macro time” constant (e.g. date), “micro time” randomized (mix up IV randomly across time). DO NOT first do all of treatment 1, then treatment 2.
- Location, environment: General setting and set-up is constant. Randomization across time can randomize setting changes across time.
- Randomization across time can also randomize unexpected events, weather changes, etc.

Randomized Block Designs

- A way of gaining experimental power by reducing variability between treatment groups on some well-defined characteristic
- NOT REQUIRED for this project
- You MAY do this if you understand it and wish to

Randomized Block Design

- Randomly assign treatments within subsamples. Main uses:
- Multiple experimenters. Sam and Robert each do 16 trials. Each will do 8 of treatment #1 and 8 of treatment #2, randomly mixing them up.
- Balance subject characteristics. Plan on studying 16 men and 16 women. *Within each gender*, each subject is randomly assigned a treatment (controlling to be sure 8 of each sex get each treatment.)
- Balance days. Experiment will be on two days. Do 16 trials on day 1, randomly assigning 8 #1 and 8 #2.

Randomized Block

Manipulated IV

Non-Manip-ulated IV (Block)

	Treatment #1	Treatment #2	Block Total
Sam (or Women)	8	8	16
Robert (or Men)	8	8	16
Treatment Total	16	16	

Cells show number of cases. Treatment totals should be equal and the number in each block should be the same for each treatment. You can sum across blocks and just analyze treatment, but more interesting also to analyze block effects.