Logic of Experiments

**Within-Subjects Design and the
Power of Randomization**

Experiment lecture/class plan

1. Introductory lecture on basic idea of experiments & randomization
2. Discussion of experiment articles: issues of measurement and hypothesis testing
3. Details on how to do the experiment
4. More detailed lectures on the logic of experimentation, controlling for extraneous variables, and drawing conclusions from experimental evidence

---

1. Experiments, Introductory Ideas

Manipulable Independent Variable

- The essential key to any experiment or quaisi-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

- Because the experimenter can manipulate the independent variable, it is possible to use that ability to manipulate to design the research to eliminate many extraneous variables
- We will return to this issue in part 4 of this lecture series

---

Between- vs. Within-Subject Designs

- Within-subjects designs: Each subject gets all treatments. Powerful design IF
  - Feasible
  - Treatments are reversible (so you can run ABABAB designs)
  - No interference between treatments
- 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?
- Logic of audit studies
- Necessity of control, concerns about reactivity

Tutoring Example

- Imagine course like statistics, chemistry
- Want to see if tutoring program helps students do better
- Independent variable = whether tutored (yes, no) NOTE: Is manipulable, can you see why?
- Dependent variable = performance on class tests

Tutoring Program

- Survey students, ask if they had a tutor
- This is final exam grade or score
- What do we conclude? Does tutoring help?
- Why is this happening?

	Grade	Score
Tutored	2.5	60
Not Tutored	3.0	75

Change Over Time for those tutored

Change scores for those tutored (within-subject before-after)
Compare score on first exam and second. Did tutoring make a difference? Are there alternate explanations?

	Midterm	Final
Score	50	60
Grade	1.7	2.3

Comparison Group

- Consider only those who got below a C (2.0) on the first test
- Before/after comparison
- Is this a valid test?

	Test 1	Test 2
Tutored	1.5	2.6
Not Tutored	1.5	1.5

Another Comparison

- One semester, no tutor. Random sample of 20. Average grade 2.6.
- Next semester, tutor. Random sample of 20. Average grade of 2.9.
- How about this test?

Matching

- **Suppose you match by age, height, sex, eye color, past GPA, ACT math score**
- **Will that solve the problem?**
- **Problem of inherent motivation**
- **No amount of matching can solve the problem of selection bias, that the subject chose the treatment group and the factors that led to that choice may be producing the outcome**

Problem of Selection Bias

- **Who goes to a tutor?**
- **Non-equivalent control groups**
- **Inherent motivation, performance (dependent variable) correlated with factors related to independent variable**

Randomization

- **Take a group of subjects, selected any way you want**
- **Actual random assignment to groups (no cheating)**
- **Equated in the statistical long run if sample is large enough. (Ideally 30 per group or more)**

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.**
- **More on this topic in part 4**

Experiment articles

Part 2

Doing the Experiment

Part 3

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
- 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 (or randomize) 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
- Important to understand this distinction, understand what a manipulated independent variable is
  - 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
- How many oceans/states etc. can be named

4

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

Manipulated IV

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

Non-Manip-ulated IV (Block)

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.

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. Each subject is randomly assigned a treatment, with control 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.
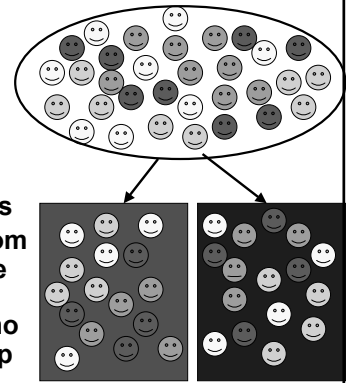
Logic of Experiments

Part 4

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

- **Take a group of subjects, and randomly divide them into treatment groups**
- **Use a true random device, eliminate any person's choice about who is in which group**

Randomization equates treatment groups

- **This works in the statistical long run**
- **Depends upon sample size**
- **Assume you have two "types" of people (I.e. males, females) that are distributed 50-50 in your subject pool**
- **NOTE: If I have only two subjects, the probability is zero. Do you see why?**

Sample size and effect of randomization

- **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?**
- **The next few slides show how sample size affects this**
- **Sample size is the number to be divided into 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

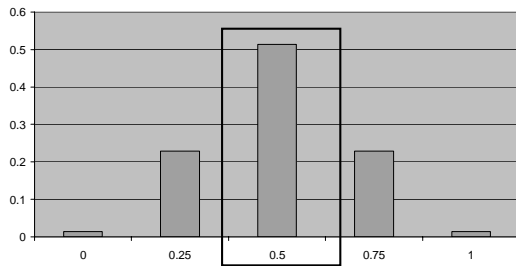| Split	Probability
100-0	1/6 = .17
50-50	4/6 = .67
0-100	1/6 = .17

Sample size 8, 4 cases per treatment

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

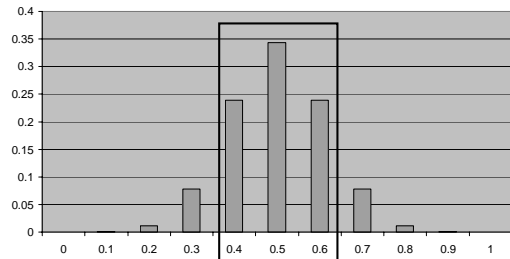| IV 1	IV2	# ways	Probability
MMMM	FFFF	1	1/70 = .014
MMMF	FFFM	16	16/70 = .229
MMFF	MMFF	36	36/70 = .514
FFFM	MMMF	16	16/70 = .229
FFFF	MMMM	1	1/70 = .014

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
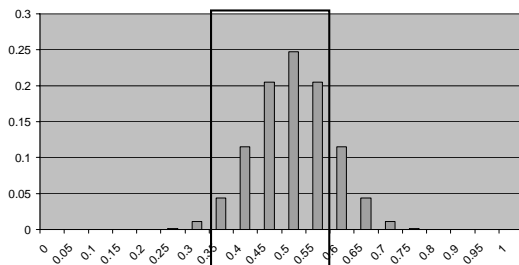
Sample size 20, 10 per group

There are 184,756 possible random divisions

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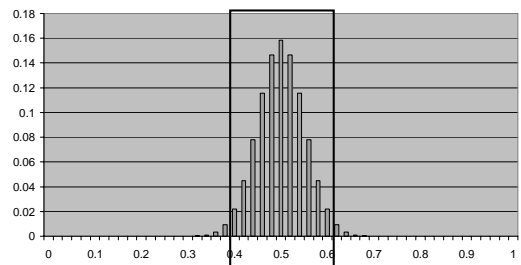Probability is .023 of a 80-20 or worse split

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**

Value of Experiments

- **If you can manipulate an independent variable, a true experiment is always the best method**
- **Best = able to isolate effect of independent variable as only cause of dependent variable**
- **Use ability to manipulate to randomize**
- **Hold constant everything not randomized**

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

Randomization in Experiments

- True experiment = random assignment of subjects to treatments (conditions) [We are doing true experiments for project]
- Quasi experiment = manipulable independent variable but no randomization
- Quasi-experimental design important in applied research where random assignment of treatments is impossible or unethical
- We are doing true experiments!!

Examples of Experiments

- Social psychololgy: e.g. Darley & Batson, Goldberg.
- 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

Constant vs. Randomization

- Constant: conditions 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.
- Randomized: more generalizable

Constant vs. Randomized: Example

- Sex of subject (subject characteristic you cannot manipulate)
- Constant = males only (or females only). Total control, but cannot generalize to the other sex
- Randomized = both sexes, let randomization equal it out. Some risk of uneven balance if sample is small. Can generalize to both sexes.

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)

Extraneous variables

- **Things that are NOT the IV that could be affecting the subject's response (DV)**
- **The statistical association between the IV and DV can be distorted only by extraneous variables that are correlated (statistically associated) with both the IV and the DV**
- **Holding constant and randomization force the statistical association to be zero between the IV and factors which are constant or randomized**
- **If you do the experiment right, almost all extraneous variables are controlled**

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**