EXPERIMENTS

Basic Design: Measure people on the dependent variable, expose them to the independent variable, measure the dependent variable again

Great control in experiments because you manipulate the IV and measure the DV and observe the whole process.

Why study experiments?
1. Experiments set the standard for controlling variation. In surveys we try to assimilate experimental control with statistical control.

2. We often do experiments with surveys -- test a question by giving different versions to different people and compare results. Example: SPOTS

3. Research jobs: do quasi experiments frequently.
Example: study at hospital with home health for infants needing IV antibiotics. Looking at parental stress for caring for infant vs. Stress in having an infant in the hospital.

Example: Poverty study

Purpose of All Research Design:
1. Systematically/scientifically answer research question
2. Control variance:
    a. Maximize experimental variance (variance of key concepts)
    b. Minimize extraneous variance (confounding variables, measurement error)
3. Experiments allow the most control over variation, and hence, theoretically, provide the best test of your hypothesis.

Example Experiment:
RQ = what is the influence of pornography on men's attitudes towards women

Experimental variance = level of pornography (hard, soft, children's film)
Extraneous variance = gender, age, race, sexuality, race, marital status

Use random assignment or matching to limit extraneous variance.

R	O_1	X_1	O_2
R	O_3	X_2	O_4
R	O_5	X_3	O_6

X = type of film watched
O = score on sexist attitudes survey

ELEMENTS OF EXPERIMENTS

Groups

1. experimental group: gets exposure to independent variable(s)
2. control group: doesn't get exposure to independent variable(s)

Variables

1. Independent Variables called factors or treatment
2. Dependent variables called outcomes

Pre and post tests

Data collected on DV before exposure to IV = pre test
Data collected on DV after exposure to IV = post test

Post-pre = change attributed to IV

Example: attitudes toward women (post-pre): stress levels of infant's parents (post-pre)

Random assignment: randomly assign participants to experimental and control groups (also called randomized study or randomization)

1. Rules out rival explanations for change in dependent variable. Eliminates systematic substantive differences between groups. Makes groups equal on variables not measured in the study (other independent variables, spurious variables)

2. Not random selection (random sampling)

3. Quasi-experiments - not random assignment. Natural groups.
a. Use this design when it is impossible or unethical to randomly assign people to exposure/experience of the independent variable. Ex: gender, race, AIDS, smoking

b. Can't rule out rival explanations with quasi-experiments because groups may not be equal on other variables. Particular problem if groups were sought because of the extreme exposure/experience on the dependent variable. For example, studying AIDS.

c. Examples: Infant study; on-line teaching

Matching: match participants in each group on an independent variable or control variable (often demographics variables).

1. Do this because:
a. You don't think random assignment will make groups equal on this particular IV

b. You can't randomly assign people to a condition: ex. race, gender, age

2. Example: studying the influence of sports on self-esteem in children. An rival explanation is height: height influences both sports participation and self-esteem. You can't randomly assign height, so you match on it.

Match kids on height. Randomly assign one in each match to the experimental group and put the other in the control group, repeat until all kids assigned.

3. Problems:
a. If participants leave a matched group the groups may become unmatched and biased

b. Works best with small sample sizes

Blinding
a. Participants can be blinded as to what study group they are in (placebo). This reduces study effects (reactivity): participants try to please researcher or displease researcher with their responses/behavior. Often still some potential for study effects because participants still know they are being studied (Hawthorne effect).

b. Data collectors can be blinded as to what study group participants are in.

Example: RQ= does education influence information seeking in the patient/physician interaction

A grad student observes interactions and codes for info-seeking. You may not want the grad student to know who is in the control or experimental group because if she knew it might bias how she perceives the interaction.

Group Size
Because there is greater control over variation in experiments, smaller n's are needed (bigger effect sizes)

Need big enough groups to detect meaningful differences

Do a power analysis to determine necessary sample size for a given difference, standard deviation(s), and alpha

Internal and External Validity

1. Internal validity = factors which influence observed values on variables, rival explanations for findings. With experiments try to maximize internal validity.

Threats to internal validity: history, maturation, mortality, testing, reactivity, instrumentation, regression (these are alternative explanations for the change from post-pre-test

Lab experiments usually high on internal validity, if groups were randomized or matched.

2. External validity = ability to generalize findings.

Threats to external validity: selection, experimental setting

Experiments usually weak on external validity because of artificial settings, artificial manipulation, selection of participants in study.

If participants were randomly selected from a population before being put into experimental and control groups, external validity increases.

Non-lab experiments usually have higher external validity as long as participants lives are not interrupted too much.

NOTATION

x= treatment, IV
o= observation/measurement on DV
R= randomized groups
---- between rows = no randomization

OVERVIEW OF TYPES OF EXPERIMENTS

Main designs: 1 DV, 1 IV
a. Pre-experimental - no control group
b. Pure experimental - Random assignment to experimental and control groups
c. Quasi experimental -experimental and control groups, no random assignment

Pre-experimental

1. One shot case study: X O
    a) No control group, no pre-test
    b) Often a group is picked that has experienced X - no manipulation of X
    c) Internal validity poor:
    no control over anything,
    anything could have caused/explain O (history, maturation, selection, mortality)

Example: DV = anxiety in statistics class
IV or (X) = teaching assistant
Start study mid-semester, so only have post anxiety levels
Won't be sure if anxiety score changed at all. Can compare anxiety scores to benchmark.

2. One group pre-post test: O_1 X O_2

No comparison group

Example: pre and post anxiety scores with statistics class.

Better than #1 but still several rival explanations for change in Os
a. History (between O_1 and O_2): longer the gap, more likely a history effect
b. Maturation (between O_1 and O_2): longer the gap, more likely a maturation effect
c. Testing
d. Study effects (reactivity)
e. Instrumentation
f. Regression (particularly if groups selected for extremity. Ex: study on the effect of education on patient info-seeking and you select people who go to the doctor a lot -- cancer pts).

3. Static Group Comparison: X O_1

O_2

Control group, but no random assignment: A group that has experienced X is compared to a group that hasn't

Example:

Fall 2002 statistics class gets teaching assistant
Spring 2003 class has no teaching assistant

Improvement over #2, but still rival explanations for change in Os:
a. Selection - no random assignment, groups not equal -- could explain change in Os. People could have sought out X . For example, students who take statistics in the Fall may differ from those who take it in the Spring.

(Example. Study on social support and AIDS victims' health. People who seek social support are a different group than those who don't seek it). Or groups could have widely different demographic representation.

b. Mortality: drop-outs bias groups even further than they may already be

True/Pure Experiments

4. Pre-post test Control group: R O_1 X O_2

R O_3 O_4

O_2-$O_{1=}$ D_1
O_4- $O_{3=}D_2$
D_{2-} D_1=D_3
D_3= effect of X

Example: Open one large section of statistics. Take all the students and randomly assign them to two sections:

Section 1: Teaching assistant
Section 2: No teaching assistant

Example: Nutrition Curriculum study (CD-ROM)

This design controls for the following internal validity threats:

a. Groups not equal/selection. Random assignment so groups are likely to be equal on other IVs and demographics (as long as groups aren't very small).

b. Testing: If no difference in O_4- O_3, than no testing effect

c. Maturation: O_4- O_3 = maturation effect. If =0, no effect.

d. History: O_4- O_3 = history effect. If =0, no effect.

e. Instrumentation: if occurs, should be likely to occur in both groups. In which case, D_2- D_1=D_3
is valid as the effect of X

f. Regression - RA makes both groups equally likely to regress, so D_2- D_1=D_3 is valid as the effect of X

Threats to internal validity that design doesn't control:
a. Instrumentation: if researcher's determine Os and are not blind
b. Reactivity
c. Mortality - control groups and experimental groups may no longer be equal.

Analysis:

R	O_1	X	O_2
R	O_3		O_4

O_2-O_1= D_1
O_4- O_3=D_2
D_2- D_1=D_3
D_3= effect of X

-Use means tests to compare D_2 and D_1 (Anova or independent samples t-test)

-Or, use ANCOVA with pre-test variable as covariate (ANCOVA is used to achieve control on some 3rd variable which couldn't be ruled out with RA or matching)

Analyses are usually simple with experiments because all the control and independent variables are built into the research design.

5. Solomon 4 group design (same s #4 but more groups)

R	O_1	X	O_2
R	O_3		O_4
R		X	O_5
R			O_6

Can assess all the rival explanations as #4, but also allows you to threats to validity.

> O_5 and O_6 allow you to remove instrumentation and testing effect from (O_2- O_1)- (O_4- O_3)
> If O_5 - O_2=0 than no instrumentation or testing effect
>
> If O_4=O_3, than no reactivity effect
>
> If O_4=O_3=O_6, than no internal threats

Analysis: same as #4, once you rule out rival explanations/threats

6. Post-test only

R		X	O_1
R			O_2

RA should make groups = on the pre-test, so pre-test is unnecessary unless you want assurance that they are =

Controls for same rival explanations as #4

Analysis: means tests on difference between O_1 and O_2

Quasi-Experimental Designs

7. Time series 1 group

O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8

Usually done with topics for which there is routine data collection: environmental issues, health care, crime data, vital statistics, economic data. Usually at the county, city, state level.

Under these conditions, strong internal validity (other than history)

Examples: unemployment, crime rates, voting numbers, poverty levels

a. Often no reactivity effects because participants don't know they are being studied (unit of analysis = city, state, etc..).

b. Testing, Mortality, Regression, Maturation = if O_1-O_2 =O_3-O_4 then no reason to believe there will be one from O_4 -O_5

c. Instrumentation: If same way of measuring Os across time, then if O_1-O_2 =O_3-O_4 then no reason to believe there will be one from O_4 -O_5

d. Selection: if same participants used across Os, then no selection effect in values of Os (example: unit of analysis = city or county or state)

Analysis:
a. Use time series methods: OLS, pooling data across years (as if you had cross-sectional data), and include a year IV. This corrects for a serial correlations. Can also sometimes do repeated measures ANOVAs

b. don't compare O_4 and O_5 mean.
c. Don't take the mean of O_1O_2 O_3O_4 and compare it to the mean of O_5O_6 O_7O_8

8. Control Group Time Series

O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8

O_9 O_{10} O_{11} O_{12} O_{13} O_{14} O_{15} O_{16}

Very strong internal validity

Example: motorcycle helmet study (compare death rates in states with and without helmet laws, and before and after law in the states with the law)

Advantages over #7: can measure history, maturation, testing, instrumentation, regression, selection, mortality effects (compare O_9-O_1, O_{10}-O_2, etc..), rather than just assume they don't exist as with #7

Analysis: same as above but use difference scores (control-experimental) as variables

External validity problem: interaction between testing and X/not X, interaction between selection and X/not X

9. Pre-test - Post-test Control group

O_1 X O_2

O_3 O_4

Example: Infant home health

Same threats and analysis as #4 (true experiment)

Strong internal validity, but regression a problem if groups picked for extremity

10. Separate sample pre-test post-test

RS	O_1		
---	---	---	---
RS		X	O_2

RS= randomly selected

Use when you can't get a pre and a post test on the same people. Ex: Military cohorts, educational cohorts.

Use one sample of a pop for the pre-test

Use another sample of a population for the post-test

Both groups get X, but post-test people don't know they will be studied when they experience X.

Examples: study on why people enlist in the military, conducted pre and post 9-11; UNCW student satisfaction pre and post new gym or library coffee bar

Strong internal and external validity:
a. No testing, selection or regression effects because of RA.

b. No Reaction effects and no interaction effects. Strong external validity.

c. Problems: history, maturation, mortality.

Analysis: $O_2 - O_1$ (means tests)

11. Separate sample pre-post test control group

RS	O_1		
RS		X	O_2
---	---	---	---
RS	O_3		
RS			O_4

Similar to #10, but you can test for history, maturation, mortality effects.

Strong internal and external validity.

Analysis: same as #4 (true experiment)

Multiple treatment (IV) designs (repeated measures designs)

1. Between-subjects designs: assign different experimental and control groups to each (like the above designs)

2. Within subjects: use same experimental and control groups, expose each to repeated IVs, or the same IV across time.

Rather than assign different people to different treatments, you expose same people to multiple treatments. Variations caused by different treatments show up in observations.

Can't do within subjects designs on variables you can't change: ex. Demographics. Can't make a person be both male/female, old/young, etc..

Problems:

a. have to be sure the effect of the 1st level of IV are gone before starting the 2nd level
b. testing effects
c. order effects

Advantage: cheaper, don't have to get new groups, use fewer cases

Example: influence on attractiveness and appropriate dress on decision to hire a woman (book)

Example: gender influences reaction to authority

DV = anger scale

Between subjects -- 4 groups, n=5 each:

Men- authority	X	O_1
Men - no authority	X	O_2
Women - authority	X	O_3
Women - no authority	X	O_4

Within subjects - 2 groups, n=5 each:

men:	X_1	O_1,	X_2	O_2
women:	X_3	O_3,	X_4	O_4

X_1 and X_3 = authority
X_2 and X_4 = no authority
O_1-O_4 = anger scale

c. Factorial: more than one IV (called a factor) in the study. Requires lots of participants.

1. Learn if each IV has an effect on the DV (main effect)
2. Learn if combinations of the IVs have an effect on the DV (interaction effects - effect of IV depends on the value of another IV).

3. If no interaction effects: effects of IV on the DV add together

Have a group for each combination of the IV

Treatment 1	Treatment 2	Factorial = 8 groups	
a	a	aa	ba
b	b	ab	bb
	c	ac	bc
	d	ad	bd

EXAMPLES of EXPERIMENTS (From Evaluating Research book)

RQ = do people use humor to cope with stress?

Asked dental patients several questions about humor to determine if they used humor to cope with stress or did not.

uses laughter:	X	O_1	O_2
	----------		
doesn't use laughter:		O_3	O_4

O_1 and O_3 = coping scale
O_2 and O_4=stress scale

Dental residents measured stress scale during dental procedure.
Patients self-rated the coping scale after the dental procedure.

Analysis: compared O_3- O_1 and O_4- O_2

Found that the laughter group rated the dental experience less stressful and had higher coping scores.

Threats/Problems:
1. Not equivalent groups -- differences may not be due to laughter. Other factors could determine stress scores (expectations going into dental appointment, experience with this doctor). Without RA can't rule out 3rd variables as explanations for findings.

2. Reactivity: being observed by dentists.

3. Testing: taking the humor scale (the original questions which placed people into the two groups) could have influenced stress levels.

RQ: Does education influence blood sugar management? Does education influence blood sugar management long term?

X_1 O_1 R: X_2 O_3

O_2

X_1=blood sugar management education
X_2=follow-up training

O_1 O_2 O_3= blood sugar levels

One group got the education. Then a random sample of them got follow-up training. Another group got no training.

Analysis: Anova - compared O_2- O_1, O_3-O_1*, O_3-O_2*,

Problems/Threats:
1. No control for 2nd phase of experiment (no O_4 data). Can't rule out history, maturation, for O_3-O_1 difference.

RQ: Do babies cry when they hear other babies crying

IV1= infant status (crier or calm baby) -- can't randomly assign to this
IV2= hear audiotape of baby crying (own or other baby's voice)

DV = duration of crying after hearing a baby cry

Factorial design: need four groups

crier baby, other baby's voice = X_1
crier baby, own voice = X_2
calm baby, other baby's voice = X_3
calm baby, own voice = X_4

O_1 O_2 O_3 O_4 = duration of crying after hearing a baby cry

R	X_1	O_1
R	X_2	O_2
R	X_3	O_3
R	X_4	O_4

Analysis: Anova - all two group comparisons. Found 2 main effects and one interaction;
1. criers cried more consistently than calm babies
2. infants who heard another baby cry cried more consistently than those who heard themselves cry
3. Crier infants cried more when heard another infant crying. Crier infants stopped crying when they heard their own voice. So the effect of hearing another baby cry, depends on whether you have a crier or a calm infant.

RQ: What is the neurological effect of alcoholism?

Quasi design: can't randomly assign people to alcoholism

O_s = learning vocabulary test

X_1 - X_3 = recovery program

Alcoholic Recovery Programs	R	X_1	O_1	O_2	O_3
	R	X_2		O_4	O_5
	R	X_3			O_6
Control Group					O_7

Design allows you to test for maturation, testing, instrumentation, and history effects

n=11 all groups, matched on age and education

covariate = initial vocabulary

Analysis
1. ANCOVA = O_3 O_5 O_6 O_7 with initial vocabulary as a covariate

2. Repeated measures ANOVA on O_1 O_2 O_3 and O_4 O_5

Found that learning not influenced by alcoholism. No neurological effects of alcoholism.

RQ = What are the social-psychological outcomes for caring for a sick infant at home versus having the infant cared for in a hospital?

X_1 O_1

O_2

X_1 = home care
O_1,O_2 = parental stress for caring for infant vs. Stress in having an infant in the hospital.

Analysis?

Threats?

Mortality - If the new (first child) parents drop out of the home health for infants study (because they are too busy to participate) than the control groups and experimental groups would no longer be equal. Could bias the estimates for the # of calls to the doctor.

Control group would have higher number of calls than the experimental group because the control group would have more first time parents in it.