Experimental Design

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

Problems in Experimental Design

True Experimental Design

- Goal: *uncover causal mechanisms*
- Primary characteristic: random assignment to sampling units
- If not random, then only *Quasi Experimental*
- Without randomization, cannot rule out some systematic biases
- Types of designs
  - *Between subject designs*: sampling units are subjected to one treatment each
  - *Within subjects designs*: sampling units receive two or more treatments

The "No Effect" Hypotheses

- Special place to the test of the hypothesis that the treatment is entirely without effect
- Reason: in a randomized experiment, this test may be performed virtually without assumptions of any kind – ie, relying only on random assignment
- Contribution of randomization is clearest when expressed in terms of the test of no effect
  - Does not mean that such tests are of greater practical importance
  - It sets randomized and non-randomized aspects in sharper contrast
- Whereas inferences in non-randomized experiments require assumptions

The “No Effect” Hypotheses

- To say a treatment has no effect is to say that that each unit would exhibit the same value of the response whether assigned to treatment or to control
- A change in response indicates the treatment has some effect
- Will discuss later various tests of the significance and the size of effects

Alternative Hypotheses

- As opposed to the null hypothesis, experimental (alternative) hypotheses take a stand
  - ★ Different treatments behave differently (two tailed prediction)
  - ★ Predict what direction the expected differences take (one tailed)

Hypotheses and Theory

- Theory is a large scale map, different areas represent general principles
- Hypotheses are like small sectional maps, focus on specific areas
- Conceptual similarities
  - range from very explicit to very vague
  - fall back on hidden assumptions, regulative principles
  - Give directions to our observations
- Some hypotheses spring from experimental observations
  - Don't know where they will go
- Others from theory
  - Conceptual hypotheses
  - Rely on previous studies, theories

Creating Hypotheses

- Defining terms
  - For observations to have value, abstractions have to be concretized
  - Two types of definitions
    - Operational: x is defined in terms of test y under conditions z
    - Theoretical: abstract constructs used
  - At some point must be operational for the experiment
  - Hypotheses are predictive statements about the expected outcome
  - They call for a test and embed a conclusion
  - Explicit statements are *de rigueur*
  - When comparisons are predicted they have to be explicated

Stating the Hypothesis

- Concomitant variation: X is a direct function of Y
- Comparative: other things being equal ….
- H0 and H1 (null and alternative) are mutually exclusive and exhaustive
  - Usually a specific H0 and a general H1
  - Try to reject H0

Hypothesis Testing Errors

- 2 types of errors: I and II
  - Type I – rejecting the null hypothesis when it is true
    - Greater psychologically important risk
    - Think there is a relationship when there is not
    - Waste time in blind alleys
  - Type II – accepting the null hypothesis when it is false
    - Deny a relationship when there is one
    - In effect, reject useful results

Variables

- Independent and dependent
  - Dependent: effect in which the researcher is interested
  - Independent: cause of the effect
  - Any event or condition can be conceptualized as either an independent of dependent variable
- Concerned about the effects of X on Y
  - Ie, the causal effects of one on the other
  - Both in the labs and in the field

Independent Variables

- No single or standard way of classifying variables
- Useful categorization (not mutually exclusive):
  - Biological
    - Eg, affects of gender in mentoring developers
  - Environmental
    - Eg, schedule pressure and fault insertion
  - Hereditary
    - Eg, IQ effects on complexity
  - Previous training and experience
    - Eg, effects of first programming languages
  - Maturity
    - Eg, age and elegance of program structures

Independent Variables

- Manipulated variables
  - If an experiment, one expects manipulation
  - Intentional and systematic variation
- Naturally occurring variables
  - Manipulated by real life experience
  - Eg, desk versus meeting inspections
  - Context: normal of exceptional conditions
- Static group variables
  - Pre-existing groups with identified characteristics:
    - Organismic variables: sex, age, weight, etc
    - Status variables: education, occupation, marital status
    - Attribute variables: diagnoses, personality traits, behaviors
  - Cannot be manipulated – but are selected to gain proper contrast groups

Independent Variables

- Analogous to experimental treatment
- When used as a dependent variable, may make inferences as to how the group acquired its characteristics
  - Eg, overweight lowers self-esteem
  - Lower self esteem causes overweightness
- Risk of causal inferences -
  - Tempting but risky
  - Dependent variable not an accurate descriptor
  - At best an association, connection, relationship, correlation
  - Example of weight/esteem experiments
    - ✓ Case 1: high calorie diet -> check esteem
      - Ethical problems
    - ✓ Case 2: overweight + low calorie -> raise esteem
      - Doesn't prove overweight, low esteem
    - ✓ Case 3: overweight + success -> lower weight
      - High esteem, low weight doesn't prove le/ow
    - ✓ Must be careful about the logic

Independent Variables

- Unidirectional paths
  - Eg, height and self-esteem – cannot switch
  - Fixed by logic of antecedents and consequences
  - Multiple variables: income, age -> truancy, discipline
    - ✓ Need 2x2 analyses
    - ✓ Question: does income discriminate truancy and discipline problems
- One-way, non-causal enabling relationships
  - Eg, income – IQ -> income
  - But not vice versa
- Two-way, sequential causation
  - Eg, success-failure and self-confidence
  - Eg, baseball players slumps, hitting streaks
- *-> Causation established by experimentation*
  - Manipulation
  - Using static variables is descriptive/relational
  - But not experimental

Independent Variables

- Establishing levels of independent variables
  - First decision: categorical or continuous
    - Eg, age is continuous, sex is categorical
    - If continuous data, whole range, dichotomous, or graduated
    - Risky as information is lost
    - May be theoretical reasons for categories
  - If hypothesis is state in categorical terms, then should be consistent
  - If a relationship, not appropriate to break into dichotomies or nominal categories if variable is continuous
    - For theoretical or rational, not statistical reasons
  - Examine how the levels of categories established
    - Should be consistent with hypothesis
    - Possible groupings: extremes, ranges of categories, median split (as in IQ)

Independent Variables

- Continuous full-range distribution
  - Sometimes linear correlations, sometimes not
  - Eg, learning (perhaps), visual acuity (not)
- Theory driven levels
  - Hypothesis stated consistently with current theory
- Strength of independent variable (magnitude of effect)
  - Extreme groupings tend to magnify effect
  - Increasing magnitude may reduce generality
  - Can more easily argue weaker to stronger (eg, stress) and have great generality
  - Levels of independent variable should match hypothesis

Dependent Variables

- Many possible flavors, literally thousands
- Eg, learning new design techniques
  - Direction of observed change
  - Amount of change
  - The ease with which change effected
  - Persistence of changes over time
- 2 general classes
  - Diffusion – fan out
    - Eg, technology insertion and adoption
  - Hierarchical variations – changes in ranking
    - Eg, changing roles in organizational structures

Practical Application

- Distinguish two kinds of control groups
  - No treatment
    - Ok for physical effects
    - Problems where belief may confound
  - Placebo
    - Rule out belief effects
- Practical decision is not easy which to use
  - Question of greatest interest
  - Experience or knowledge of the general area
  - Easy to make mistakes in a new area

Basic Designs

- Design 1
  - One shot case study: X O
  - H- M- I(NR) S-
    - Deficient in terms of any reasonable controls
      - ✓ History may be alternative explanation
      - ✓ Maturation not controlled for
      - ✓ May be changes in instruments or judges
      - ✓ Unknown state of participants
    - Instrumentation not a factor: no pre-measurement
- Design 2
  - One group pretest: O X O
  - Slight improvement, but no comparison

Basic Designs

- Design 3 – Solomon Design
  - True experimental, 4 group
    - I R O X O
    - II R X O
    - III R O O
    - IV R O
  - H+ M+ I+ S+
    - All well controlled for

Basic Designs

- Solomon provides elegant illustration of logic of control
  - Pretest performance scores in I & III to estimate pretest scores in II & IV
    - Requires a leap of faith even if scores the same
    - Even if differ greatly, II & IV could be equal to the mean of I & III
  - Use estimated pre-test scores to enrich factorial analysis of variance of post test scores
  - Tells us if any confounding of pre-test and treatment

Basic Designs

- Pre/post test effects:
  - I+ II- III+ IV-
- Experimental treatment effects:
  - I+ II+ III- IV-
- Pretest & X sensitization:
  - I+ II- III- IV-
  - Pretest sensitization = $(\bar{Y}_I - \bar{Y}_{III}) - (\bar{Y}_{II} - \bar{Y}_{IV})$
- Extraneous effects:
  - I+ II+ III+ IV+

Basic Designs

- External Validity
  - All 3 suffer from the possible confounding of selection and treatment
  - Design 3 - Solomon:
    - controls for confounding of treatment and pre-test sensitization
  - Design 4:
    - I R O X O III R O O
    - IV: H+ M+ I+ S+
    - Deficient in pre-test sensitization – eg, problem in attitude change or learning experiments
  - Design 5
    - II R X O IV R O
    - IV: H+ M+ I+ S+
    - Avoids pretest sensitization issues

Basic Designs

- Within subjects designs
  - Each subject receives all treatments in turn
  - Useful in SWE/CS – repeated measures design
  - Advantages:
    - Same number of subjects used more effectively
    - Each sampling unit serves as its own control
    - Can examine relationships longitudinally
  - Difficulties:
    - Sensitization problems
      - ✓ Learning etc
    - Order of treatments may produce differences in successive measures
  - Another threat to IV in longitudinal studies
    - Regression towards mean
      - ✓ When linear relationship is imperfect
      - ✓ Eg, overweight people appear to lose weight, low IQs appear to become brighter
      - ✓ Observed when variables consists of the same measure taken at two points in time and the correlation $r < 1$

Basic Designs

- Solve threat by standard *Z* score
  - A raw score from which the sample mean has been subtracted and the difference then divided by the standard deviation
  - Regression equation: $Z_Y = r_{XY} Z_X$
    - The estimated score of *Y* is predicted from the *XY* correlation r times the standard score of *X*
    - If there is a perfect correlation, the *Z* scores will be equivalent; otherwise not if $r < 1$