

PLS 506
Mark T. Imperial, Ph.D.

Lecture Notes:
Data Collection

- Research
 - Undertaken for three broad purposes
 - Explore a phenomenon to become familiar with it and to gain insight
 - Describe a particular community, group, situation as completely, precisely, and accurately as possible
 - Examine and formally test relationships among variables
 - *Exploratory study* is conducted when little is known about a subject. The goal is both description and understanding
 - *Descriptive study* is conducted to describe some phenomenon. It is a fact finding exercise and focuses on describing the dimensions of a phenomena systematically.
 - *Explanatory study* tests relationships to formally seek the answers to problems and hypotheses. It focuses on describing the nature of specific relationships between specific variables.
- Stages of Social Research
 - Selection and formulation of the research problem
 - Preparation of the research design
 - Overall plan and framework for the investigation
 - Unit of analysis, variables to be observed and controlled, the nature of their relationships, how the variables should be operationalized and measured, how these data will be collected, what will be tested, how will these data be analyzed,
 - Measurement: have to link specific concepts to empirically observable events
 - Sampling: how many units selected and how to go about choosing them
 - Data Collection: Variety of methods including
 - Archival data sources like agency records
 - Surveys
 - Focus groups
 - Interviews
 - Trained observer ratings/direct observation
 - Expert judgments
 - Role playing (e.g., secret shoppers)
 - Data processing: data must be processed or transformed for analysis
 - Data analysis and interpretation: Need to determine how these data will be analyzed
- *Unit of Analysis*
 - The entities (objects or events) under study
 - People, roles, positions, social groupings (families, organizations, cities, etc.) or artifacts such as books, documents, buildings
- *Ecological fallacy*
 - When relationships between properties of groups or geographic areas are used to make inferences about individual behaviors of people within groups or areas

- Example: foreign born persons commit more crimes than native born persons because the crime rate is higher in areas with greater proportions of foreigners. It is false because we don't know who committed the crimes – foreigners might have been the victims
- Types of variables
 - *Dependent variable (Y)*: the one the researcher is interested in explaining or predicting
 - *Independent variable (X)*: the variables that explain variation in the dependent variable.
 - *Antecedent variables*: occur prior in time to both dependent and independent variables
 - *Intervening (mediating) variables*: are an effect of the independent variable and a cause of the dependent variable.
 - *Extraneous variables*: not part of the explanatory set. Control variables are held constant and kept from varying during the course of observation or analysis. Uncontrolled variables will vary.
 - *Quantitative variables*: its values or categories consist of numbers and differences between categories can be expressed numerically. Example: Income
 - *Qualitative variables*: have discrete categories usually designated by words, labels, or concepts. Example: Gender:
- Relationships between variables
 - If our goal is discovery or description, research findings will depend in large part on what relationships are anticipated and which ones are uncovered
 - Relationships or associations depend on: (1) if one variable changes then another changes in some way; (2) if one variable does not change, then the other does not change.
 - Three properties of relationships:
 - (1) direction;
 - (2) nature of association (linear, curvilinear);
 - (3) and strength of association.
 - *Positive or direct relationship*: Increase in one variable is accompanied by an increase in another while a decrease in the same variable is accompanied by a decrease in the other variable.
 - *Negative or inverse relationship*: An increase in one variable is accompanied by a decrease in the other variable.
 - *Linear relationships*: rates of change are constant.
 - *Curvilinear relationship*: rates of change are not constant
 - *Strength of relationship*: indicates how closely the variation in the variables is connected. Correlation coefficients are a measure of the strength and direction of linear relationships
 - *Causality*: For one variable X to cause another Y there must be some statistical association between X & Y. X must also precede Y in time. Direction of influence must be consistent with hypothesized relationship so that independent variable causes the dependent variable not vice versa. Difficult or impossible to prove. So we often look for disconfirming evidence.

- Spurious relationships: A spurious relationship may exist when the dependent and independent variables are related to some extraneous variable. This can cause a spurious association. Good research designs attempt to control for all extraneous factors that might cause spurious relations in order to eliminate rival hypotheses.
- Hypotheses
 - A hypothesis is an expected but unconfirmed relationship between two or more variables
 - Stated as
 - *If-then (conditional) statements*: If you are X, then Y will occur
 - *Mathematical statements*: $Y = f(X)$ or “Y is a function of X”
 - *Continuous statements*: “the greater the X, the greater (or lesser) the Y”
 - *Difference statements*: One variable differs in terms of the categories of another variable. People with high incomes are more educated than people with low incomes.
 - Null hypothesis is used to disconfirm the hypothesis