

Sociology 357

Methods of Sociological Inquiry

Lectures Notes #2
Basic Concepts

Units of Analysis

- The objects we study
- People
- Families
- Cities
- Newspaper articles
- Classes (in school)

Variables

- Dimensions or aspects of units of analysis which vary. Variables **MUST** vary.
- Formal definition of a variable is a set of **exhaustive** and **mutually exclusive** categories.
 - Every unit of analysis must fall into exactly one category of a variable.
- Variables are defined by researchers

Examples of Variables

UNIT OF ANALYSIS	VARIABLE	CATEGORIES OF VARIABLE
individual	income	exact income to nearest dollar or categories, e.g. < \$10,000 \$10,000 - \$24,999 \$25,000 - \$34,999
individual	eye color	blue, brown, green, hazel, etc
Census tract	Average income	Mean to nearest dollar
organization	sex composition (% female)	percentage to nearest whole percent, or categorize: <20% female 20-50% female 51-80% female >80% female

Level of Measurement

- **Nominal.** Exhaustive & mutually exclusive categories. E.g. Eye color, major, sex/gender.
- **Ordinal.** Nominal + ranks. Course grade (A, AB, B etc.); course level (E, I, A)
- **Interval.** Ordinal plus meaningful metric so distance between 1 & 2 = distance between 23 & 24. Few examples: temperature scales.
- **Ratio.** Interval + true zero. Height in inches, income in dollars, number people enrolled in a class.

Qualitative/ Quantitative

- **Qualitative** = Nominal + ordinal = qualitative. Can do frequencies, percentages, proportions, mode.
- **Quantitative** = Interval + ratio Can do qualitative + means, standard deviations, correlations, all other statistics.
- Very few statistics especially for ordinal. Ordinal variables with 5+ categories can usually be assigned numbers and treated as interval.

Propositions

- A **proposition** is a statement about variables.
- A **univariate** proposition is a statement about one variable at a time. "Most UW students drink beer at least once a week."
Variable: frequency of beer drinking.
UOA: individual, UW students. Statement: "most" drink once a week or more.

Bivariate Proposition

- A **proposition** is a statement about variables.
- A **bivariate** proposition is a statement about the **relation** between two variables. "Males drink beer more often than females."
Variables: 1) sex, 2) frequency of beer drinking. Statement: gives relation between them.

Multivariate Proposition

- A **proposition** is a statement about variables.
- A **multivariate** proposition states a **complex relation** among three or more variables. "Among non-depressed students, males drink beer more often than females, but among clinically depressed students, males and females drink beer equally often." Variables: 1) sex, 2) frequency of beer drinking, 3) whether clinically depressed or not.

Hypothesis

- A **hypothesis** is a type of proposition.
- Some use proposition and hypothesis as synonyms.
- I use hypothesis to mean the proposition being tested in a particular research project. This is the most common usage.
- Some use hypothesis to mean a proposition whose truth is uncertain.
- (Stern uses hypothesis for the bivariate finding of a project, even if it is an after-the-fact result.)

General Form of a Hypothesis

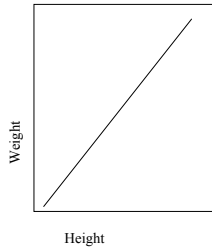
- Conceptual: For population P in condition C, independent variable X causes dependent variable Y.
- Operational: For sample p in condition c, independent variable x has a statistical association with dependent variable y.

Qualitative Relations

- Used with qualitative variables.
- Need to be stated in words, listing which categories of one variable have more or fewer units of analysis in each category of the other variable.
- Ex: Blacks are more likely to be Democrats than whites are. Variables: race, party choice.
- Percentages

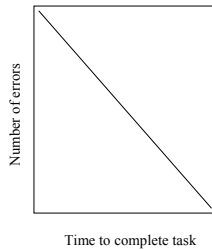
Quantitative Relationships - 1

- Between quantitative variables
- **Positive** = when variable one is greater, the other tends to be too. Height is positively correlated with weight. On the average, taller people weigh more



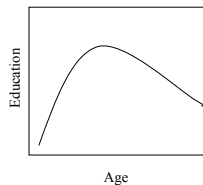
Quantitative Relationships - 2

- **Negative** = when one variable is greater, the other tends to be smaller. The speed of performing a task is negatively related to accuracy. On the average, the faster you work, the more mistakes you make.



Quantitative Relations - 3

- A **curvilinear** relation can be any non-linear relation, but is especially a relationship that is first positive then negative, or vice versa.



There is a curvilinear relation between age and education in the US: education rises with age but, because of historical increases in education rates, older adults have less education than younger adults.

Operationalization=Measurement

- To operationalize a variable is to say how you will measure it
- To measure a variable is to use specific observational or operational procedures
- The operationalization of a concept is the same thing as its measure or measurement
- This has two parts

Operationalization Part I

- First, the procedures for collecting data (e.g. observe, ask questions)
- Question: “How often have you smoked marijuana?” vs “Have you every smoked marijuana?”
- Observation of fidgets: Take motion picture, count frames in which position changes vs. observe face to face, count number of times hands touch head.

Operationalization Part II

- Exact distinctions among categories of variables within a procedure
- If counting, how you tell the beginning and end of countable things
- If distinguishing among types of actions or characteristics, must develop rules for an exhaustive and mutually exclusive set of different types

Operational Variables are Created by Researchers

- No measured variable is “natural.” All have to be created by the decisions of the researcher. (But some are easier to operationalize than others.)
- Researcher makes sure the categories are exhaustive and mutually exclusive
- Researcher decides how precise to be

Precision vs. Accuracy

- A more precise variable makes finer distinctions.
 - Height in inches instead of feet.
 - Shadings of eye color grey-blue, sky blue, deep blue, violet-blue, blue-green, pure green, yellow-green, light brown, dark brown etc. instead of broad groups blue, brown, hazel
- Accuracy is correct classification into the category.
- Tradeoff between precision and accuracy. Harder to be accurate with finer distinctions.

Range

- Categories must be exhaustive, so must encompass the full range the subjects exhibit
- “Other,” “over \$100,000,” and “not applicable” are ways to make a variable exhaustive
- Categories should be defined so only a few subjects end up in the residual category

Indicators

- Indicators are indirect measures of variables
- Not really operationalization of the thing itself, but something known to be highly correlated with it.
- e.g. “Where there’s smoke, there’s fire.” Smoke is an indicator of fire.
- Church attendance is an indicator of religiosity. You might attend regularly for crass, non-religious reasons (or not attend due to practical constraints), but there is a high correlation between frequency of church attendance and subjective religiosity.
- Indicators are important for concepts which are difficult to measure directly.
