Chapter 1 – Research Inquiry

Required Reading

Babbie, Chapter 1: Human Inquiry and Science

Terms to Define

Variable vs. Attribute

Inductive vs. Deductive Theory

Nomothetic vs. Ideographic explanation

Logico-empirical

Social Regularities

Causal and Probabilistic Reasoning

Prediction vs. Understanding

Replication

Epistemology vs. Methodology

Questions for Discussion

1. How do tradition and authority hinder inquiry? How do they support it? Give examples.

2. What are the main requirements of science? How does astrology measure up to these requirements?

3. Explain why the errors in personal human inquiry that Babbie describes are so common.

Chapter 2 and McLeod & Pan: Paradigms & Concepts

Required Reading

Babbie: Ch. 2 Paradigms, Theory, and Social Research
McLeod & Pan: Concept Explication and Theory Construction

Terms to Define

Concept Explication

Dimension

Indicator

Meaning Analysis

Theory

Research Hypothesis

Scientific Concept/Construct

Conceptual Definition

Dependent vs. Independent Variable

Abstractness vs. Clarity in Meaning

Correlation

Spurious Relationship

Questions for Discussion

1. **Give a short conceptual definition of a familiar communication concept, then an operational definition that is consistent with your conceptual definition.**

2. **How do scientific concepts differ from everyday concepts. What are the scientific standards that everyday concepts do not meet?**

3. **What are the steps necessary for concept explication?**

4. **Why is research a cyclical process?**

5. **Explain the steps of theory construction.**

Chapter 3 & 4: Research Ethics and Design

Required Reading

Babbie, Chapter 3: The Ethics and Politics of Research
Babbie, Chapter 4: Research Design

Terms to Define

Voluntary Participation

Anonymity and Confidentiality

Deception and Debriefing

Institutional Review Boards

Cohort Studies

Cross-sectional Studies

Longitudinal Studies

Panel Studies

Ecological Fallacy

Reductionism

Social Artifacts

Triangulation

Units of Analysis

Description vs. Explanation vs. Exploration

Questions for Discussion

1. Discuss the reasons why a social scientist would do an exploratory study, a descriptive study, and an explanatory study. Select a research topic and show how this topic would be addressed in terms of each of the three purposes of research.

2. Compare the ecological fallacy with reductionism in terms of similarities and differences. Give examples of each. Present a case for one being a more serious error than the other.

3. Briefly compare and contrast the different roles that researchers may play in field research, that is, complete participant, participant observer, observer participant, and complete observer.

Chapter 5: Reliability and Validity

Required Reading

Babbie, Chapter 5: Conceptualization, Operationalization, and Measurement

Terms to Define

Nominal Measures

Ordinal Measures

Interval Measures

Ratio Measures

Reliability

Validity

Face Validity vs. Content Validity

Criterion-Related Validity

Construct Validity

Discriminant validity (divergent validity)

Questions for Discussion

1. Provide the reasoning for the following statement: "A research instrument can be reliable but not valid."

2. What are the four levels of measurement? Describe and give an example of each.

3. Differentiate reliability from validity. Which is more important? Why?

4. What is the difference between the criterion-related validity and construct validity? In what type of research would one be more important than the other?

5. Discuss possible threats to the internal validity of an experiment.

6. What are the criteria for inferring causality? How do experiments versus surveys differ in meeting these criteria?

Chapters 6 & 7: Measurement/Sampling

Required Reading

Babbie, Chapter 6: Indexes, Scales, and Typologies

Babbie, Chapter 7: The Logic of Sampling

Measurement

Terms to Define

External Validity

Index/Scale

Item Analysis

Thurstone Scale

Likert Scale

Semantic Differential

Guttman Scale

Typology

Exhaustive vs. Mutually Exclusive

Loaded Question vs. Leading Question

Questions for Discussion

1. What are the advantages and disadvantages of using self-administered questionnaires compared to interviews? How is the presence of an interviewer a potential disadvantage in survey research? An advantage?.

2. Compare and contrast the following scaling procedures: Likert, Thurstone, and Guttman. For each, note purpose, advantages, disadvantages, and procedures.

3. Discuss the advantages and disadvantages of open (free-answer) vs. closed (fixed alternative) response questions.

4. What problems should a researcher be aware of when deciding to use the response "don't know?"

Sampling

Terms to Define

Representativeness

Confidence Interval vs. Confidence Level

Sampling Interval vs. Sampling Ratio

Population vs. Sample

Parameter vs. Statistic

Purposive Sampling

Simple Random Sampling vs. Systematic Sampling

Stratified vs. Quota Sampling

Sampling Frame

Sampling Distribution

Convenience Sampling

Questions for Discussion

1. What are the factors that influence the decision of sample size?

2. Define sampling error. Can we totally eliminate sampling error in sampling? How, if we can? If the answer is no, why? How is sampling error related to population size, sample size, and the degree of precision?

3. When would you use a simple random sample (SRS)? When would it be better to use systematic sampling?

Chapter 8: Experimental Research

Required Reading

Babbie, Chaper 8: Experiments

Terms to Define

Pretest

Posttest

Double-blind experiment

Internal versus external validity

Control group

Randomization versus matching

Natural experiment

Chapter 9: Survey Research

<u>Required Reading</u>

Babbie chapter 9: Survey Research

<u>Terms to Define</u>

Questionnaire

Respondent

Open-Ended Questions

Closed-Ended Questions

Bias

Contingency Question

Response Rate

Interview

Probe

Questions for Discussion

1. What are advantages of the experimental method compared to surveys? What are advantages of surveys over experimental methods

2. How is causality established and what are required criteria?

3. Why is randomization so important and preferred to matching?

4. Why is a low response rate of concern in survey research? Why is response bias of concern? Which is a more serious threat to survey results and generalization—low response rates or response biases? Why?

5. Compare the strengths and weaknesses of surveys and experiments.

6. Why is it important to record the respondent's answers to open-ended questions exactly? What kinds of errors can result from the failure to record answers correctly? How would incorrect recording affect internal and external validity? What about reliability?

Monge & Contractor.: Social Network Analysis

Required Reading

Monge & Contractor, *Theories of Communication Networks, Chapters 1 & 2*

Terms to Define

Node

Relationship

Tie

In-degree/Out-degree

Betweenness

Centrality

Network Size

Network Heterogeneity

Network Density

Questions for Discussion

1. What are advantages of social network analysis for understanding phenomena?

2. Why is social network analysis so well suited to studying social media?

3. How can network analysis reveal indirect relationship between actors in a social network?

Chapters 10 & 11: Unobtrusive Measures /Content Analysis

Required Reading

Babbie, Chapter 10: Qualitative Field Research

Babbie, Chapter 11: Content Analysis

Terms to Define

Coding

Content Analysis

Latent Content

Manifest Content

Unobtrusive Measures

Analysis of Existing Statistics

Ideal Types

Historical/Comparative Analysis

Ethnography

Ethnomethodology

Focus group

Grounded theory

Questions for Discussion

1. Compare and contrast content analysis, analysis of existing statistics, and qualitative field research in terms of purposes, strengths, and weaknesses.

2. Explain why units of analysis are of particular concern in content analysis.

3. Explain why social science researchers are interested in both manifest and latent content. Show how these two types of content are both similar and different.

Chapter 14: Qualitative data analysis & Hypothesis Testing

Required Reading

Babbie, Chapter 14: Quantitative data analysis

Terms to Define

Null Hypothesis

Research (Directional) Hypothesis

Type I Error

Type II Error

Significance Level

Statistical Power

Concept mapping

Grounded theory method

Semiotics

Variable versus case-oriented analysis

Questions for Discussion

1. Why are concepts of Type I and Type II error so central to hypothesis testing and the goals of the research?

2. Explain the difference between directional and non-directional hypotheses?

3. Explain how approaches like grounded theory or semiotics can be used alongside or in place of conventional hypothesis testing research.

Chapter 16: Data Analysis & Statistics

Required Reading

Babbie, Chapter 16: Statistical Analyses

Terms to Define

Multivariate Analysis

Univariate Analysis

Mean

Median

Mode

Dispersion

Standard Deviation

Continuous Variables vs. Categorical Variable

Discrete Variable

Contingency Table

Control Variable

Questions for Discussion

1. Compare univariate and bivariate analyses in terms of definition, purpose, advantages, and limitations.

2. Explain the logic in constructing and interpreting bivariate tables. Why should you percentage within categories of the independent variable and then compare across categories?

3. Why do we use controls? What logic should we use to govern how we choose control variables? Is it possible to use too many controls?

Statistics

Terms to Define

Descriptive Statistics

ANOVA

Factor Analysis

Inferential Statistics

Statistical Significance

Measures of Association

Pearson's r

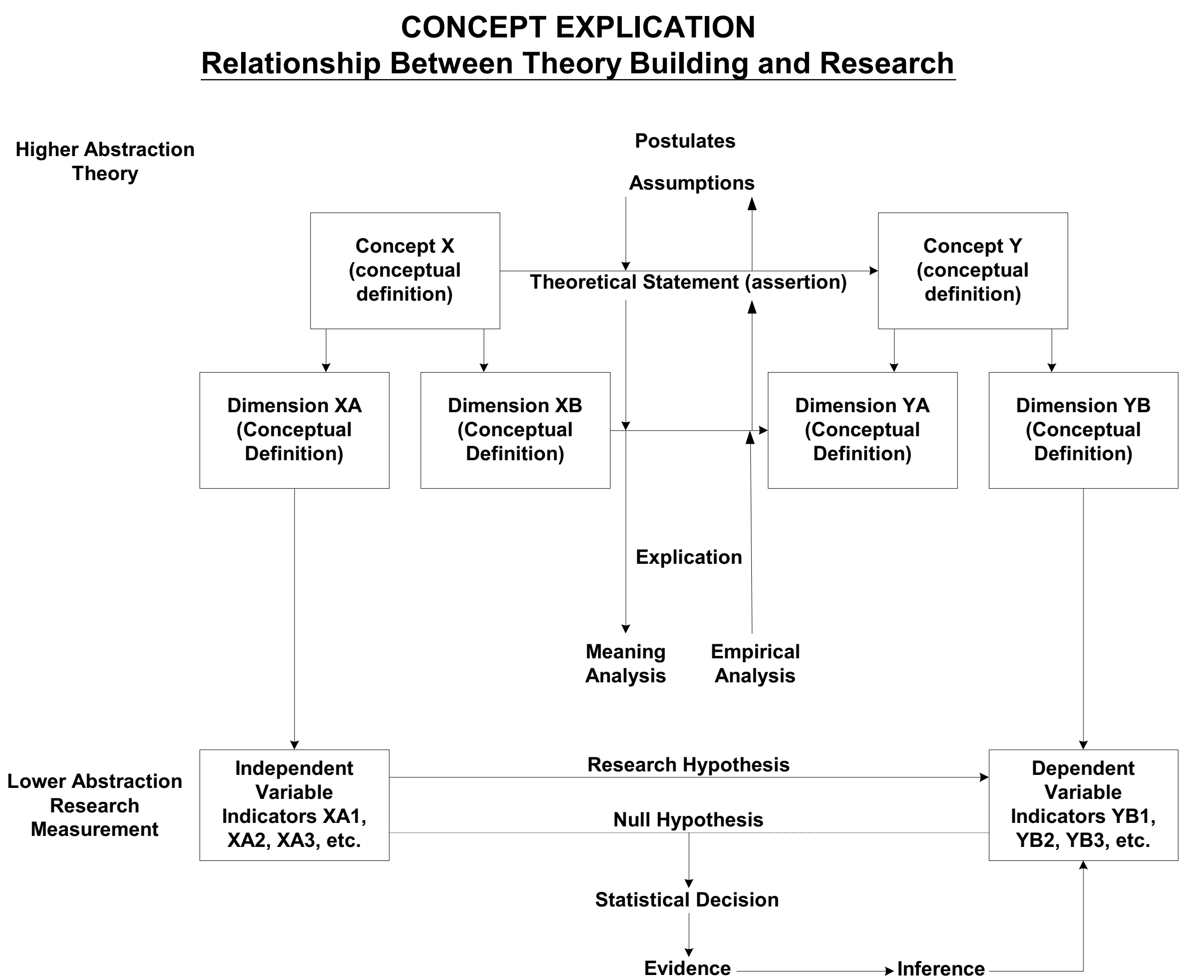
Chi-Square

Questions for Discussion

1. Explain the two main uses of Factor Analysis? What purpose do they serve for research?

2. Explain the appropriate situation for the use of ANOVA? How about Chi-Square? Are there any similarities between the two?

CONCEPT EXPLICATION:



CONCEPT EXPLICATION GLOSSARY

Abstractness: A characteristic of scientific concepts. It refers to the degree to which a concept is distanced from some specific events or phenomena observed at some particular time and place, or to the range of objects or subjects whose characteristic is represented by the concept.

Antecedent: Events or phenomena that occur before some other events or phenomena and impact on the latter. They are sometimes called antecedent variables.

Clarity in meaning: A characteristic of scientific concepts. It refers to the degree to which its intended meaning is clearly stated and conveyed. It also refers to the degree to which clarity is achieved in a specific use of a concept.

Class concept: A concept that denotes some collection of objects or subject by summarizing the communalities among them.

Concept: A term or phrase used to identify an object or subject, or to represent a common characteristic of a class of object or subjects. They are the building blocks of thought and theory.

Concept explication: The process by which abstract concepts are linked to their real world variations so that they can be observed by appropriate methods. Logically, it involves both deductive and inductive reasoning. It can be further divided into meaning analysis and empirical analysis.

Conceptual definition (theoretical, constitutive): Involves verbal descriptions of the essential properties that are to be included in the intended meaning of a concept. In research practice, it often specifies various dimensions of a concept.

Consequence: Events or phenomena that occur after the events or phenomena under consideration and are affected by them. They are often called consequent variables, or more simply, effects.

Dependent variable: In either theoretical statements or research hypotheses, it is the variable concept whose fluctuation (variation) is to be explained. It is assumed to be affected by the independent variable to which it is linked by a relational statement.

Dimension (sub-concept): Distinguishable components of a more abstract concept that have higher levels of coherence than has the concept. Various dimensions of a concept may have different antecedents and consequences and thus theoretical statements must be stated for the constituent dimensions rather than for the more abstract concept.

Empirical analysis: Statistical and logical procedures involved in using research data to evaluate the quality of conceptual and operational definitions of a concept.

Epistemological relationship: The conceptual linkage between an abstract concept and its real world referents (indicators), which could be a specific object or subject, a class of objects or subjects, or observable variations of a characteristic across a class of objects or subjects.

Explanation: A set of logically organized statements specifying why and how some observed events or phenomena have occurred. Because it lays out the specific conditions under which conceptually similar phenomena occur, it can be used to derive various specific predictions.

Extension: The opposite of intention. It refers to the conceptually constructed linkages between a concept and its real world referents.

Independent variable: In either theoretical statements or research hypotheses, it is variable concept used to explain fluctuation (variation) in the dependent variable. It is assumed to affect or influence the dependent variable to which it is linked by a relational statement.

Indicator: Concrete observable behaviors (etc.) indicating a concept or a dimension. It is concrete in that it can be directly or indirectly recorded by some research technique. Its epistemological relationship with the concept is that it is assumed to be invariant within some specified range of time, space and domain.

Manipulation: Procedure in experimental research where different inductions (e.g., message forms, instructions) are given to randomly assigned sets of subjects to estimate differences in response in some measured or observed dependent variable.

Meaning analysis: Logical and epistemological procedures used in developing conceptual and operational definitions of a concept. Along with empirical analysis, it comprises the iterative process of concept explication.

Measurement: Assignment of values (numbers or symbols) to units of observation according to rules. The rules are those specified in the operational definitions and reflect standards used by scientists or scholars in a given field. Measurement includes nominal or categorical assignment as well as more powerful forms: ordinal, interval and ratio measurement.

Meta-concept: A composite concept representing multiple characteristics of a class of object or subjects, often filled with affect and/or values. It can be decomposed into several concepts (e.g., democracy into a class concept representing a type of political system, or into a variable concept indicating the degree to which democratic ideals are practiced.

Null hypothesis: A statistical statement that asserts that there is no relationship between the variables in the research hypothesis, or more precisely than any relationship or difference shown is simply due to chance or sampling error.

Observation: The recording of data from observational units pertinent to a given variable and compatible with its operational definition. Used very broadly and well as specifically applied. Specifically, it is a research method where an observer records data either as a non-participant (e.g., scores group interaction) or as participant in studying a group over time while playing some role in their lives. More broadly, it refers to the various forms of obtaining data from observational units: direct observation, interviews, self-administered questionnaires, etc.

Observational unit: The collection of singular objects that we actually observe. They may or may not correspond to the units of analysis in any research project. Where they do not correspond, the data obtained from the observational units may be aggregated to provide measures of variables for the units of analysis.

Operational definition: Procedures by which a concept is to be observed (as in participant observation), measured (as in sample surveys), or manipulated (as in experiments). It details the rules, specific steps, equipment, instruments, and scales involved in measuring (etc.) a concept. All three types of operational procedures can be generally called either observation or measurement, using these two terms in their broadest sense.

Operationalizability: A characteristic of scientific concepts. It refers to the degree to which the linkage between the concept and its real world referents can be specified and its variation across a class of objects or subjects can be observed in the real world.

Operationism: A doctrine about scientific research which holds that the only fruitful and valid meaning of a concept is its measurement or operational procedure, nothing more and nothing less.

Precision: A characteristic of scientific concepts. It refers to the degree to which the concept is precisely measured and its intended meanings is agreed upon by communicators (scientists) and receivers (consumers of scientific research or policy makers).

Prediction: A statement about the likelihood that some specific event or phenomenon will occur in some designated future time.

Premise (assumption): Propositions that supply the reasons for a theoretical statement. Premise is a term from symbolic logic; two or more premises are needed to draw a logical conclusion. Assumption is its synonym when applied more generally in theory and research. Assumptions should be testable with empirical evidence and often they are based on a considerable body of evidence from previous research. Assumptions are distinguished from postulates, which are more abstract untestable statements about human nature, world views value statements, etc.

Primitive term: Word or phrase whose meaning is widely shared and incapable of further definition except by using synonyms. They are used to define other theoretical terms in a theory.

Referent: Conceptual implied and empirically observable counterparts of a concept in the real world.

Relational concept: A concept that makes connections between other forms of concepts. The connections can be comparative (e.g., larger than, as long as) associative (e.g., positively related, negatively related), or causal (e.g., leads to, caused by).

Research hypothesis: Assertions about the relationship of two or more variables stated as concrete operational definitions. Their relationship is predicted from the logic contained in the corresponding theoretical statement and in the explication of the variables contained in the statement. A research hypothesis is stated before the empirical evidence is examined. The assertion describes the nature and/or the magnitude of the relationship.

Scientific concept (construct): A concept that has been consciously invented, constructed or adopted for purposes of theory building and research. It should be clearly defined and may be broken down into less abstract (more specific) dimensions. It should meet four standards: abstractness, clarity in meaning, operationalizability and precision.

Singular concept: A concept that denotes some particular object or subject.

Theoretical statements: Assertions about the relationship of two or more relatively abstract variable concepts or their dimensions connected by a relational statement. They should be clear as to the units of analysis of concern and the conditions under which the statement should hold.

Theory: An organized explanation of some recurrent phenomena of research interest. It is a collection of (1) theoretical statements about the relationship between two or more variable concepts or their dimensions within a domain or collection of units of analysis under specified conditions; (2) the premises or assumptions providing the reasoning behind the theoretical statement; (3) two or more variable concepts and their conceptual and operational definitions; and (4) specific research hypotheses connecting operational definitions of the concepts included in the theory.

Unit of analysis: The collection of singular concept objects collected into and described as a class concept. Variation in the units of analysis is the focus of any research project. Units of analysis commonly used in mass communication research include: individuals, media organizations, media systems of nations, communities, families, newspapers etc.

Variable concept: A concept that distinguishes objects or subjects in terms of the degree to which they possess some designated characteristic (e.g., level of education, age, television viewing hours per day).

Working definition: A convenient operational procedure linking a concept to its real world variations for a specific research purpose. It does not necessarily imply a conceptual meaning of the concept and, thus, may not be generalizable beyond the specific research purpose.

RELIABILITY AND VALIDITY

Sociologist James A. Quinn states that the tasks of scientific method are related directly or indirectly to the study of similarities of various kinds of objects or events. One of the tasks of scientific method is that of classifying objects or events into categories and of describing the similar characteristics of members of each type. A second task is that of comparing variations in two or more characteristics of the members of a category. Indeed, it is the discovery, formulation, and testing of generalizations about the relations among selected variables that constitute the central task of scientific method.

Fundamental to the performance of these tasks is a system of measurement. S.S. Stevens defines measurement as "the assignment of numerals to objects or events according to rules." This definition incorporates a number of important distinctions. It implies that if rules can be set up, it is theoretically possible to measure anything. Further, measurement is only as good as the rules that direct its application. The "goodness" of the rules reflects on the *reliability* and *validity* of the measurement--two concepts which we will discuss further later in this lab. Another aspect of definition given by Stevens is the use of the term numeral rather than number. A numeral is a symbol and has no quantitative meaning unless the researcher supplies it through the use of rules. The researcher sets up the criteria by which objects or events are distinguished from one another and also the weights, if any, which are to be assigned to these distinctions. This results in a scale. We will save the discussion of the various scales and levels of measurement till next week. In this lab, our discussion will be focusing on the two fundamental criteria of measurement, i.e., reliability and validity.

The basic difference between these two criteria is that they deal with different aspects of measurement. This difference can be summarized by two different sets of questions asked when applying the two criteria:

Reliability:

a. Will the measure employed repeatedly on the same individuals yield similar results? (stability)

b. Will the measure employed by different investigators yield similar results? (equivalence)

c. Will a set of different operational definitions of the same concept employed on the same individuals, using the same data-collecting technique, yield a highly correlated result? Or, will all items of the measure be internally consistent? (homogeneity)

Validity:

a. Does the measure employed really measure the theoretical concept (variable)?

EXAMPLE: GENERAL APPROACHES TO RELIABILITY/VALIDITY OF MEASURES

1. Concept: "Exposure to Televised News"

2. Definition: the amount of time spent watching televised news programs

3. Indicators:

    a. frequency of watching morning news
    b. frequency of watching national news at 5:30 p.m.
    c. frequency of watching local news
    d. frequency of watching television news magazine & interview programs

4. Index:

Design an eleven-point scale, where zero means "never watch at all," one means "rarely watch" and ten "watch all the time." Apply the eleven-point scale to each of the four indicators by asking people to indicate how often they watch each of the above TV news programs.

Combining responses to the four indicators/or survey questions according to certain rules, we obtain an index of "exposure to televised news program," because we think it measures TV news exposure as we defined it above. A sum score of the index or scale is calculated for each subject, which ranges from 0 (never watch any TV news programs) to 40 (watch all types of TV news program all the time). Now, based on the empirical data, we can assess the reliability and validity of our scale.

DETERMINING RELIABILITY

1. Stability (Test-Retest Correlation)

Synonyms for reliability include: dependability, stability, consistency (Kerlinger, 1986). Test-retest correlation provides an indication of stability over time. For example, if we asked the respondents in our sample the four questions once in this September and again in November, we can examine whether the two waves of the same measures yield similar results.

2. Equivalence

We want to know the extent to which different investigators using the same instrument to measure the same individuals at the same time yield consistent results. Equivalence may also be estimated by measuring the same concepts with different instruments, for example, survey questionnaire and official records, on the same sample, which is known as multiple-forms reliability.

3. Homogeneity (Internal Consistency)

We have three ways to check the internal consistency of the index.

a) Split-half correlation. We could split the index of "exposure to televised news" in half so that there are two groups of two questions, and see if the two sub-scales are highly correlated. That is, do people who score high on the first half also score high on the second half?

b) Average inter-item correlation. We also can determine internal consistency for each question on the index. If the index is homogeneous, each question should be highly correlated with the other three questions.

c) Average item-total correlation. We could correlate each question with the total score of the TV news exposure index to examine the internal consistency of items. This gives us an idea of the contribution of each item to the reliability of the index.

Another approach to the evaluation of reliability is to examine the relative absence of random measurement error in a measuring instrument. Random measurement errors can be indexed by a measure of variability of individual item scores around the mean index score. Thus, an instrument which has a large measure of variability should be less reliable than the one having smaller variability measure.

DETERMINING VALIDITY

1. Criterion (Pragmatic) Validity

Based on different time frames used, two kinds of criterion-related validity can be differentiated.

a) Concurrent validity. The measures should distinguish individuals --whether one would be good for a job, or whether someone wouldn't. For example, say a political candidate needs more campaign workers; she could use a test to determine who would be effective campaign workers. She develops a test and administers it to people who are working for her right now. She then checks to see whether people who score high on her test are the same people who have been shown to be the best campaign workers now. If this is the case, she has established the concurrent validity of the test.

b) Predictive validity. In this case our political candidate could use the index to predict who would become good campaign workers in the future. Say, she runs an ad in the paper for part-time campaign workers. She asks them all to come in for an interview and to take the test. She hires them all, and later checks to see if those who are the best campaign workers are also the ones who did best on the test. If this is true, she has established the predictive validity of the test and only needs to hire those who score high on her test. (Incidentally, criticisms of standardized tests such as GRE, SAT, etc. are often based on the lack of predictive validity of these tests).

2. Construct Validity

Three types of evidence can be obtained for the purpose of construct validity, depending on the research problem.

a) Convergent validity. Evidence that the same concept measured in different ways yields similar results. In this case, you could include two different tests. For example:

1. You could place people on meters on respondent’s television sets to record the time that people spend with news programs. Then, this record can be compared with survey results of “exposure to televised news”; or

2. You could send someone to observe respondent’s television use at their home, and compare the observation results with your survey results.

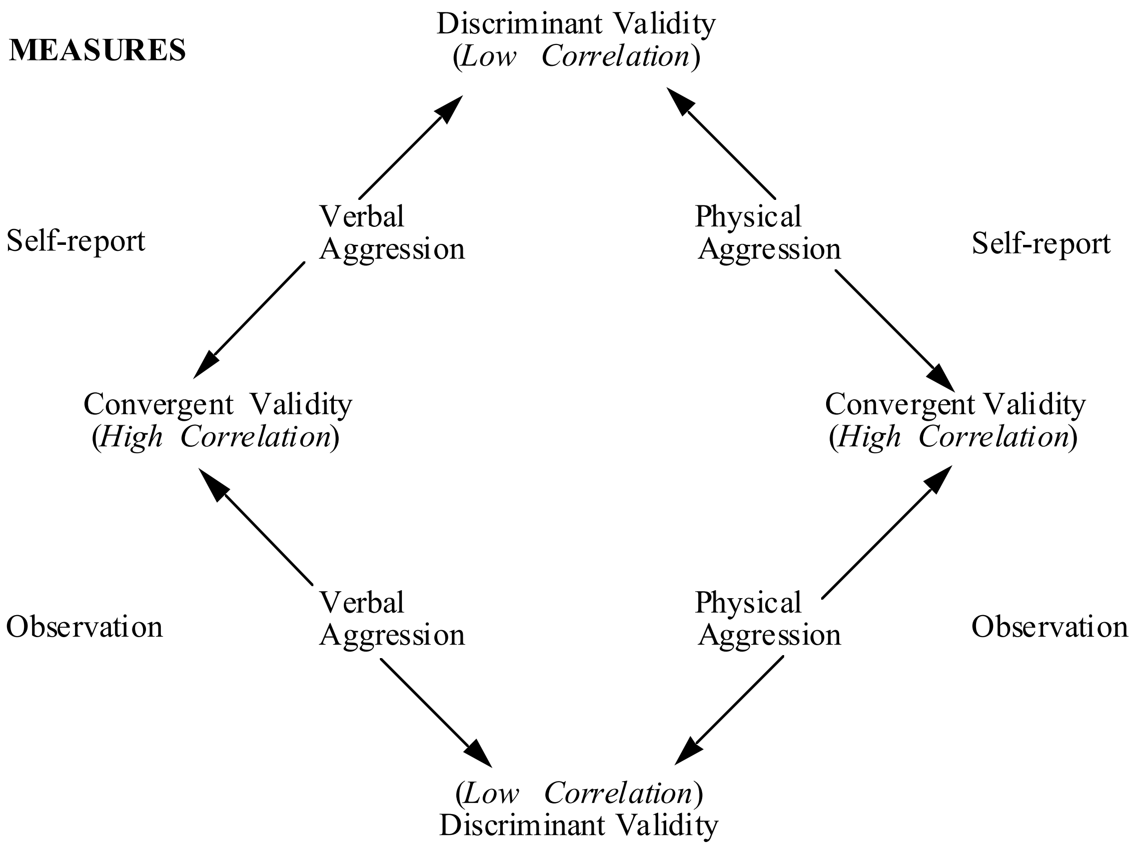
b) Discriminant validity. Evidence that one concept is different from other closely related concepts. So, in the example of TV news exposure, you could include measures of exposure to TV entertainment programs and determine if they differ from TV news exposure measures. In this case, the measures of exposure to TV news should not related highly to measures of exposure to TV entertainment programs.

Convergent Validity: Where different measures of the same concept yield similar results. Here we used self-report versus observation (different measures). Yet, these two measures should yield similar results since they were to measure verbal (or physical) aggression. The results of verbal aggression from the two measures should be highly correlated.

Discriminant Validity: Evidence that the concept as measured can be differentiated from other concepts. Our theory says that physical aggression and verbal aggression are different behaviors. In this case, the correlations should be low between questions asked that dealt with verbal aggression and questions asked that dealt with physical aggression in the self-report measure.

Example: Convergent/Discriminant Validity

Theoretical Statement: Physical violence in television leads to physical aggression.



c) Hypothesis-testing. Evidence that a research hypothesis about the relationship between the measured concept (variable) and other concept (variable), derived from a theory, is supported. In the case of physical aggression and television viewing, for example, there is a social learning theory stating how violent behavior can be learned from observing and modeling televised physical violence.

From this theory we derive a hypothesis stating a positive correlation between physical aggression and the amount of televised physical violence viewing, then, can be derived. If the evidence collected supports the hypothesis, we can conclude a high degree of construct validity in the measurements of physical aggression <u>and</u> viewing of televised physical violence since the two theoretical concepts are measured and examined in the hypothesis-testing process.

3. Face Validity

The researchers will look at the items and agree that the test is a valid measure of the concept being measured just on the face of it. That is, we evaluate whether each of the measuring items matches any given conceptual domain of the concept.

4. Content Validity

Content validity regards the representativeness or sampling adequacy of the content of a measuring instrument. Content validity is always guided by a judgment: Is the content of the measure representative of the universe of content of the concept being measured (Kerlinger, 1986)?

Although both face validation and content validation of a measurement is judgmental, the criterion for judgment is different. While the belonging of each item to the concept being measured is to be determined in the evaluation of face validity, content validation determines whether any left-out item should be included in the measurement for its representativeness of the concept.

An example may clarify the distinction. Now, the task here is to determine content validity of a survey measure of "political participation." First, we may specify all the aspects/or dimensions of this concept. Then, we may take the measurement apart to see if all of these dimensions are represented on the test (e.g., the questionnaire). For example:

POLITICAL PARTICIPATION

Dimensions	Behavior: Expressing own viewpoint	Behavior: Learning other's viewpoint	Cognitions
Indicators	Political activity	Viewing broadcasts	Interest in politics
	Voting registration	Discuss with family/friends	Party affiliation
	Voted in past	Reading campaign materials	Political knowledge
	Membership in organizations		

Have we left out any dimensions? If we are not representing all the major dimensions of the concept, we've got low validity. We won't be measuring some aspects of the concept. Some people will probably get different "scores" on the political participation test than they should, since we haven't measured some of the things we need to. You can think of the domain of the concept "political participation" as a universe consisting of different aspects (dimensions). The measures of the concept are a sample from the universe. The question dealt with in content validity is whether the sample (measurement) is representative enough to cover the whole universe of the concept domain.

Presented in the following are two tables outlining the different ways of establishing reliability and validity. TABLE 4-1 shows that, to establish any form of reliability, one needs two or more independent observations on the same people. As we may realize later, the more independent observations we have on measurement of a concept taken with different points of time or forms, the more freedom we gain to establish reliability.

TABLE 4-1

TYPES OF RELIABILITY

		Time dimension	
		Multiple-Time-Point Study	Single-Time-Point Study
Forms	Multiple	Equivalence Stability	Equivalence
	Single	Stability	
Items	Multiple	Homogeneity Stability	Homogeneity
	Single	Stability	

TABLE 4-2 shows different types of validity and three criteria which distinguish them. The three criteria are where to start the validation, the evidence and criteria for establishing validity. As you may see, construct validity is the most demanding in that both theory and empirical data are required in the process of validation. Nonetheless, it is the most valuable in theory construction.

TABLE 4-2

TYPES OF VALIDITY

Validity types	Where to Start	Evidence	Criteria
Judgmental (Pre-Data)			
Face Validity	Indicator	Judgmental	What's there
Content Validity	Concept	Judgmental	What's not there
Data-Based (Post-Data)			
Criterion-Related Validity 1. Concurrent 2. Predictive	Criterion Group 1. criterion manifesting currently 2. criterion occurring in the future	Empirical	*Empirical Criterion* Prediction
Construct Validity	Theory	Empirical	*Theoretical Criterion* Convergent Discriminant Hypothesis-testing

SAMPLING

Sampling is the foundation upon which almost all research is built. Having an understanding of sampling procedure is an absolute necessity not just to conduct research but to be able to interpret the research done by others. While this laboratory will not cover all possible sampling variations, it will present the basic steps involved in all sampling and detail one particular method. Briefly, there are three steps to draw a sample:

1. determine precisely ***the universe*** to be studied;
2. specify the desired ***degree of precision*** with which the sample must reflect the true state of the population sampled from; and
3. select an appropriate ***sampling method***.

1. **Determining the Universe**. Determining the universe specifies a theoretical population which will be examined. It also establishes the theoretical boundaries for the population, which will guide the construction of sampling frame. The ***sampling frame*** is the actual list of all the units (elements) comprising the statistical universe. For instance, a researcher may want to determine the approximate number of years of formal education attained by adults residing in the United States. The universe is all adults, defined as any individual eighteen years of age or older, residing in the United States. The researcher is now faced with the problem of constructing the actual list of all American adults from which to draw a sample. One possible solution would be for the researcher to set up a sampling frame (element list) from voter registration files. While this solution might work, there are serious problems connected with its use. Aliens residing in the U.S. are not permitted to vote and therefore would not show up on the list. Persons moving a lot and persons with low educational attainment would also be underrepresented. These may be important omissions which could bias the study's results. But for some other research purposes, for example, voting preferences among all voters, such sampling frame may present no problem whatsoever. The point is that whenever constructing a sampling frame one must always keep the theoretical universe which the sampling frame should approximate carefully in mind. When compromises must be made, the ramifications of those compromises must be carefully considered.

2. **Degree of Precision**. The specification of the degree of precision for a sample usually takes the form of a statement such as "Our sample mean will fall within two standard deviations of the true population mean 95% of times if we draw the sample 100 times." In this statement, the degree of risk in accepting the sample as an accurate reflection of the true population has been specified. The researcher is sure that her sample mean will fall within a particular tolerance range 95% of the times. In this case, the tolerance range is two standard deviations in a <u>sampling distribution</u>. A ***standard deviation*** is a statistical term indicating a standardized distance from a mean of a distribution. The standard deviation of a sampling distribution is often called the ***standard error***, which indicates a standardized distance from the mean estimated by all possible samples of the same size drawn from the same population. To say a score has fallen within two standard deviations of the true population mean is the same as saying a score has fallen within 47.5 percentage points on either side of the population mean. In other words, we can consider a range of scores within which each sample mean falls. This range consists of

all possible means if we draw samples from the same population over and over again. Two standard deviations from the true population mean represents the range within which 95% of the sample means fall, 47.5% above, and 47.5% below the true population mean.

An example may further clarify the notion of degree of precision. Let's return to the case of estimating the average years of formal education attained by adults residing in the States. Say, we drew a probability sample of 1,500 and obtained an estimate of 14.5 years. This number may differ from the true average years in the population, designated as X. But we know that there is 95% chance that this estimate, 14.5 Years, falls within a particular range. This range can be expressed as X $\pm$ 2 standard errors. And range is conceived as the degree of precision of the estimate.

Now, we know the degree of precision is determined by the size of standard error. And, the size of a standard error is primarily decided by sample size. The larger the sample size, the smaller the standard error is. In other words, the larger the sample, the greater precision is in estimating the true population mean from the sample mean. Since the size of a sample generally represents time and cost, it is important for a researcher to know exactly how precise his estimate of the true population mean must be before investing a large amount of resources to draw a huge sample. For a pilot study on consumer acceptance of a new dish washing liquid, the degree of precision, and thereby the sample size, could probably be less than for a FDA commissioned study concerning a possible lethal food additive.

One note of restraint here. While increases in sample size do yield increases in accuracy, the trade-off is not perfect. That is, doubling your sample size does not guarantee twice the precision. In fact, there is a point where even vast increases in sample size add very little to precision.

3. **Sampling Procedure.** While there are several procedures available, this lab will only concern itself with the use of ***simple random sampling (SRS).*** You should be, however, familiar with the concepts that lie behind the use of other methods.

SRS is a form of probability sampling. The key concept in SRS is that every element in the defined population has an ***equal*** chance of being selected into the sample. Whereas, its variants, i.e., other forms of probability sampling, differ from SRS in that every member of a population has a ***known and non-zero*** chance of being selected into the sample. When a sample is so selected, probability theory can be applied to the data yielded by the sample. This is an extremely important point since probability theory provides the base for all inferential statistics.

As its name implies, SRS is simply the random selection of elements from the sampling frame. Following are the steps involved in drawing a sample using SRS.

1. Assign *one and only one* number to each element in the sampling frame. Begin with the number one and number each element consecutively. If there were ten

elements in the sampling frame -- Chicago, New York, Los Angeles, Baltimore, Miami, New Orleans, San Francisco, Portland, Washington, and Madison -- each would be assigned a number in consecution: Chicago-1; New York-2; Los Angeles-3; Baltimore-4; Miami-5; and so on until all elements were assigned.

2. Enter a table of random numbers (refer to Singleton *et al.*, pp. 139-140) and select the correct number of elements for the sample. For example, to randomly select a sample of three from the above sampling frame of ten cities, one would enter the table of random numbers randomly and read two digits. The numbers smaller than or equal to ten indicate the corresponding elements in the sampling frame to be selected. The procedure continues till the desired sample size is achieved. The number of digits to be read together is determined by the total number of elements in the sampling frame. For example, to select a sample of three used car dealers from a sampling frame of 100 dealers, three digits would be read at a time. We should enter the table of random numbers randomly and read the table either horizontally across the columns or vertically down the rows. For instance, let's say the random entry in this case happens to be the second row down from the upper left-hand corner and we decide to read across the columns. Since the sampling frame consists of three digits, and we decide to read the last three digits of each set of the numbers (or, you could read the first three digits). The sample in this case would consist of the dealers who were assigned the numbers of 037, 080, and 019. If the sampling frame consisted of 1,000 elements, then four digits would be read at a time and so on.

The next issue we are about to discuss concerns the reliability of *estimates from samples*. Estimates of population parameters from sample statistics are subject to a type of error known as ***sampling error***. The notion of sampling error is demonstrated graphically in FIGURE 1 through 6. These figures illustrate the mean values of six different variables: education, age, political knowledge, TV public affairs content exposure, newspaper public affairs content exposure and income. All data are from the 1985 MCRC survey of Dane County conducted by the J658 class.

Each figure contains 21 bars, each representing the mean value of a variable as estimated from a given sample. The shaded center bar, designated by the letter P (for population), represents the mean from the entire group of 512 respondents. The bars to the left of center represent the 10 means as estimated from samples of size 25. The 10 values to the right are estimated from samples of size 105. Note the estimated values for the latter are more stable, that is, have fewer extreme values and smaller ranges. (***Range*** is defined as: the highest score minus the lowest score.)

In addition, there are three types of probability distributions with which we will be typically concerned:

a. a <u>population</u> (or parent) distribution;
b. a <u>sample</u> distribution; and
c. a <u>sampling</u> distribution.

Population distribution

As the term suggests, the population distribution is the probability distribution of the variable in the ***population***. When describing this distribution, we use ***parameters***, i.e., the mean, mode, median, standard deviation, and variance.

Suppose we ask the question, "How many days a week do you read a newspaper? " What is the population distribution of this variable in Madison? Say, the total population in Madison is 312,434; everyone is interviewed; the mean is 3.6 days per week.

Sample distribution

Suppose we did a survey of 500 people in Madison and came up with a sample distribution with a mean of 3.5, which is slightly different from the population mean as might be expected due to error.

The frequency distribution and its corresponding percentages are shown in TABLE 2. The first two columns on the left show the population distribution and the two columns on the right show the sample distribution. A graph of the sample distribution is shown in GRAPH 8-1. As you may see, the distribution is fairly close to a ***normal distribution***. (Notice that our mean is around 3.5. Most of the responses are between 3.0 and 4.0; some are 2.0 and 5.0; and a few are 1.0 and 6.0; very few are 0 or 7.0.)

To describe the distribution in the ***sample***, we refer to values such as the *mean, mode, median, standard deviation, and variance* as ***statistics***. Sample statistics are like population parameters, except they describe samples instead of populations.

Sampling distribution

Now, we find 35 other researchers who have conducted surveys this year in Madison with samples of 500 each using this same question. How is this sample statistic distributed over samples? How are the means, for example, of all these 35 separate samples distributed?

Treating this statistic as a random variable under consideration, we could graph the probability distribution of means. We look at these sample means and find that the average (the mean of the means) of the distribution is around the population mean. By plotting the 35 sample means, we can see this to be true.

1	survey had a mean of	2.8
2	surveys had means of	3.0
3	surveys had means of	3.4
6	surveys had means of	3.5

11	surveys had means of	3.6
7	surveys had means of	3.7
4	surveys had means of	4.1
1	survey had a mean of	4.5

The distribution of some sample statistic over samples is called a ***sampling distribution***. Yet, we can talk about many types of sampling distributions. In this case we considered the sampling distribution of means of samples with size 500. We could just as well consider the way in which sample standard deviations are distributed over samples, or sample variance, etc.

SUMMARY

We can infer from this *sampling distribution* that the actual population mean is very close to 3.6 although we may never know what it actually is. This property of sampling distributions, however, is used as a basis in statistical inference.

A ***frequency distribution*** shows the actual number of each of the attributes of a variable (e.g., days per week reading a newspaper. This may differ from the *normal distribution* (normal curve).

Population and sampling distributions, however, refer to *content*: A ***population distribution*** is the *actual distribution* (which we rarely ever know but instead infer from the sample); while a *sampling distribution* is the distribution of a statistic (such as a mean) across *many samples*--a normal distribution about a population parameter.