SACRAMENTO STATE
Leadership begins here.

Gathering Research Data

Stephen E. Brock, Ph.D., NCSP
California State University, Sacramento

1

What is Quantitative "Data"?

- The different types of data gathered as part of an empirical study are referred to as *variables*.
- All variables have at **least two and often** more values or scores.
- Variables can be either **categorical** (e.g., eye color, gender) or **quantitative** (e.g., rankings, test scores).
- Variables take the form of at least one of four **Scales of Measurement**.
- Different scales require different types of statistical analysis.
- QUESTION: What is data in a qualitative study?

2

Scales of Measurement

Scale	Properties	Examples
Nominal **Qualitative** (Categorical) Variables	Qualitative **categories**. Observations sorted into categories by principle of **equivalence**. Scale categories differ from one another only in a qualitative (not quantitative) sense.	Eye color Gender Ethnicity Type of school ADHD v no ADHD
Ordinal **Quantitative** Variables	Observations are **ranked** in order of magnitude. Ranks express a "**greater than**" relationship. No implication about how much greater.	Ordinal 1 = Tallest 6'7" 2 = 6' 3 = 5'11" 3 = 5'11" 5 = 5'8"

Scales of Measurement

Scale	Properties	Examples
Interval **Quantitative** Variables	Numerical value indicates **order AND meaningfully reflects relative distances**. A given interval between measures has the same meaning **at any point** in the scale.	Educational Tests
Ratio **Quantitative** Variables	Scale has all properties of an interval scale, **AND** has an **absolute zero point.**	Length Weight

4

Scales of Measurement

◆ Family <u>Income</u> and Student Reading <u>Test Scores</u>
  ▪ How is one quantitative (*ratio*) variable **related** to another quantitative (*interval*) variable?
    ▪ Correlation study

◆ **<u>Gender</u>** and Student Reading **<u>Test Scores</u>**
  ▪ How is one categorical (*nominal*) variable **related** to another quantitative (*interval*) variable?
    ▪ Correlation study

5

Scales of Measurement

◆ **<u>Homework vs. Longer Classes</u>** and Math **<u>test scores</u>**
  ▪ How does one categorical (*nominal*) variable **affect** another quantitative (*interval*) variable?
    ▪ Ex-Post Facto or Experimental Study.

◆ **<u>ADHD</u>** (Y/N) and Reading comprehension **<u>test score</u>**
  ▪ How does one categorical (*nominal*) variable **affect** another quantitative (*interval*) variable?
    ▪ Ex-Post Facto study

6

Activity

- ◆ State a research question
  - Identify the **scale of measurement** used in addressing the research question.

- ◆ Identify the association OR cause and effect relationship between variables.
  - Identify the **type of study**

7

Group Comparison Variables

- ◆ Independent Variable (IV; the **cause**):
  - The variable hypothesized to have a given effect.

- ◆ Dependent Variable (DV; the **effect**):
  - The variable used to measure the hypothesized effect.
    - ◆ AKA "Dependent Measure"

8

Which variable is the IV?
Which variable is the DV?

- ◆ [IV] [Homework vs. Longer Classes] and [Math Test Scores] [DV]
  - How does one Categorical (nominal) variable **affect** another quantitative (interval) variable?
- ◆ [IV] [ADHD] and [Reading comprehension] [DV]
  - How does one Categorical (nominal) variable **affect** another quantitative (interval) variable?

9

Methods of Data Collection

- Useful in quantifying both the IV & DV
  - Standardized measures
    - e.g., published tests
    - These is portfolio assignment
      - Why was this type of data emphasized by making it a portfolio assignment?
  - Experimental measures
    - i.e., measures developed by the researcher.
  - Coding
    - of observations and records.

10

Methods of Data Collection

- Standardized measures
  - e.g., published tests.
  - Use of will make data collection much easier

11

Types of Measuring Instruments

- Cognitive Tests (what people know and how they think).
  - Achievement Tests
  - Aptitude Tests (e.g., IQ tests)

- Affective Tests (what people believe, feel, and perceive).
  - Attitude Scales
  - Interest Inventories
  - Personality Inventories

12

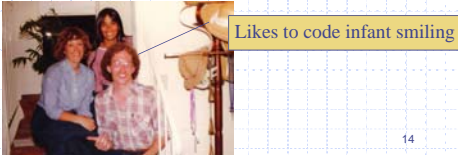## Methods of Data Collection

◆ Experimental measures
- i.e., measures developed by the researcher.
- e.g., reading comprehension test (See supplemental handout on my webpage)

13

Methods of Data Collection

◆ Coding
- of observations and records.
- e.g., systematic behavior observation techniques (see subsequent slides)
- e.g., infant smiling code (see supplemental handout on my webpage)

Likes to code infant smiling

14

Systematic Observation:
Data Collection

◆ Event Frequency Data
- **Definition**: Number of occurrences of behavior that has a clear beginning and end, measured over a specified time period.
- **Example of behaviors measured**: A punch; runs from room; shouts out response, words read per minute, hand raises, number of problems completed, eye blinks, questions answered correctly, self-injurious acts with a clear beginning and ending.
- **Advantages:** Easy to record. A small golf counter is often used to collect this type of data.

Reference: Sulzer-Araroff, B., & Mayer, G. R. (1991). *A guide to selecting behavior recording techniques. Behavior Analysis for Lasting change.* New York: Holt, Rinehart & Winston. 15

Systematic Observation:
Data Collection

◆ Event Frequency x Activity Data
- Activity Scatter Plot
- Helps to identify if the frequency if a given behavior is greater during specific activities.

Activity	Frequency
Art	✓
Transition	✓✓
Math	✓✓✓✓
L.A.	✓✓✓✓✓✓
Reading	✓✓✓✓✓✓✓✓✓✓✓✓✓✓
Free time	

16

Systematic Observation:
Data Collection

◆ Event Frequency x Time Data
- Time Scatter Plot
- Helps to identify if the frequency of a given behavior is greater during specific times of the day.

Time	Frequency
8:00-8:15	✓
8:15-8:30	✓✓
8:30-8:45	✓✓✓✓
8:45-9:00	✓✓✓✓✓✓
9:00-9:15	✓✓✓✓✓✓✓✓✓✓✓✓✓✓
9:15-9:30	

17

Systematic Observation:
Data Collection

◆ Event Frequency Data

Behavioral event to be counted

Date	Frequency	Notes

18

Systematic Observation:
Data Collection

◆ Duration Data

- **Definition**: Length of time from beginning to end of a response. If a behavior may last several minutes and/or does not occur very frequently, then this is a preferred data source.
- **Example of behaviors measured**: Temper tantrums, time spent on task, amount of time out of seat, length of time to sit down following teacher request to do so, length of a temper tantrum, or any behaviors where duration is an important variable.
- **Disadvantages:** Required the use of a clock or stop watch.

Reference: Sulzer-Araroff, B., & Mayer, G. R. (1991). *A guide to selecting behavior recording techniques. Behavior Analysis for Lasting change.* New York: Holt, Rinehart & Winston. 19

Systematic Observation:
Data Collection

◆ Duration Data

Behavioral event to be counted and timed

DATE:	DATE:	DATE:	DATE:	DATE:
Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:
Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:
Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:
Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:
Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:	Start: Stop: Duration:

20

Systematic Observation:
Data Collection

◆ Permanent Product Data

- **Definition**: The enduring outcome of the behavior.
- **Example of behaviors measured**: Number of problems or number of assignments completed, windows broken. Activities with discrete, countable segments.
- **Advantages:** Reliability, Can be collected after the fact in some cases (e.g., by looking a teacher grade books).

Reference: Sulzer-Araroff, B., & Mayer, G. R. (1991). *A guide to selecting behavior recording techniques. Behavior Analysis for Lasting change.* New York: Holt, Rinehart & Winston. 21

Systematic Observation:
Data Collection

◆ Permanent Product Data

Behavioral outcome (or product) to be counted		
Data Collection		
Date	Frequency	Notes

22

Systematic Observation:
Data Collection

◆ Interval Data
- **Definition**: Number of time *intervals in which the behavior occurs at least once.* Total observation time is divided into equal intervals and noting the behavior's presence or absence during that time. If the behavior occurs frequently (at least once every 15 minutes), then this is the preferred data source.
- **Example of behaviors measured**: Thumb sucking, on/off-task, gestures, stereotypical behavior
- **Advantages:** Records behaviors that are not clearly discrete (not have real clear beginnings and endings).

Reference: Sulzer-Araroff, B., & Mayer, G. R. (1991). *A guide to selecting behavior recording techniques. Behavior Analysis for Lasting change.* New York: Holt, Rinehart & Winston.   23

Systematic Observation:
Time Sampling Techniques

- **Whole-interval time sampling.** Records the response when displayed *throughout the entire interval*. Can be used to measure on-task behavior. Tends to underestimate occurrences of behavior. Useful when it is important to know that the behavior has not been interrupted.
- **Partial-interval time sampling.** Records the response when a single instance is displayed *at any time during the interval*. Can be used to measure swearing or bizarre gestures. Tends to overestimate occurrences of behavior. Used to record behaviors that are fleeting.
- **Momentary-interval time sampling.** Records the response if it is displayed *at the end for a specific interval*. Can be used to measure in-seat behavior or frequent stereotypic behavior. Useful to record behaviors that are apt to persist for a while.

Reference: Sulzer-Araroff, B., & Mayer, G. R. (1991). *A guide to selecting behavior recording techniques. Behavior Analysis for Lasting change.* New York: Holt, Rinehart & Winston.   24

Systematic Observation:
Data Collection

◆ Interval Data (whole, partial, momentary)

Behavioral event to be counted

Interval:	DATE:	DATE:	DATE:	DATE:	DATE:
8:00					
8:15					
8:45					
9:00					
9:15					
9:30					
9:45					
10:00					
10:15					
10:30					
10:45					
11:00					
11:15					
11:30					
11:45					
12:00					

25

How do you quantify attitudes, interests, beliefs, feelings, & traits?

◆ Likert Scales
- e.g., Agree = 3, Undecided = 2, Disagree = 1

◆ Semantic Differential Scales
- Use bipolar adjectives
- e.g., Necessary __ __ __ __ __ Unnecessary

◆ Rating Scales
- e.g., Always = 3, Sometimes =2, Never = 1

◆ All require a self-report along a continuum of choice.

26

Evaluating Tests (Rudner, 1994)

◆ In addition to evaluating their use in empirical research, why else is it important for us to attend to this guidance?

◆ Test coverage and use.
- There must be a clear statement of recommended uses and a description of the population for which the test is intended.

◆ Validation and norming samples.
- The samples used for test validation and norming must be of adequate size and must be sufficiently representative to substantiate validity statements, to establish appropriate norms, and to support conclusions regarding the use of the instrument for the intended purpose.

27

Evaluating Tests: Reliability

- ◆ The degree to which a test **consistently** measures a variable.
  - The degree to which you can count on it
  - How dependable it is.
- ◆ A valid test is always reliable, but a reliable test is not always valid.
- ◆ In your own words, why is it important for a researcher to know that his or her test is reliable?

28

Evaluating Tests: Reliability

- ◆ Stability (test-retest reliability)
  - changes over time.
- ◆ Equivalence (alternate form reliability)
  - Similarity of two versions of the same test.
- ◆ Internal Consistency
  - Similarity of items within a test.
- ◆ Scorer/Rater
  - Agreement of independent scores/raters.

29

Standard Error of Measure (SEM)

- ◆ Allows us to estimate how much difference there probably is between a person's **obtained** and **true** scores. The size of the difference is a function of the tests reliability. Big differences indicate low reliability.
- ◆ Reporting scores as falling within a given range (confidence intervals) takes SEM into account.

30

Standard Error of Measure (SEM)

- ◆ A reflection of the measurement error that exists in most of the "tests" used in the social sciences.
- ◆ Some measures have more error than others.
- ◆ The greater the error (SEM) the lower the reliability.
- ◆ What if a measure being used in an educational research study does not have reliability statistics?

31

Evaluating Tests: Validity

- ◆ A test consistently measures what it is supposed to measure
  - Permits appropriate interpretations
  - Valid for specific purposes and populations.
  - A matter of degree.

- ◆ Why is the validity of a test (or DM) important to the educational researcher?

32

Evaluating Tests: Validity

- ◆ Content Validity
  - The test measures the intended content area.
  - Includes both item (item relevance to content area) and sampling (sample of total content area) validity.
  - Determined by expert judgment.
    - ◆ For example:
      - The content validity of a science test would be determined by a group of experienced science teachers

33

Evaluating Tests: Validity

◆ Criterion-Related Validity

- A test (the predictor) correlates with a second measure (the criterion)

- Includes concurrent validity.
  - ◆ Both measures administered in same time frame.
  - ◆ How well measure reflects current functioning.
    - For example:
      - Correlation between the experimenter's 7th science test results and student grades given by their 7th grade science teacher.

34

Evaluating Tests: Validity

◆ Criterion-Related Validity (continued)

- A test (the predictor) correlates with a second measure (the criterion)

- Also includes predictive validity
  - ◆ Both tests administered at different times
  - ◆ How well measure predicts future performance
    - For example:
      - Correlation between the experimenter's 7th grade science test results and student grades given by 8th grade science teacher.

35

Evaluating Tests

◆ Construct Validity
- The test scores reflects the construct it is intended to measure
- Requires a series of studies
  - ◆ Including content & criterion-related validity studies
- Most important form of validity
- Does the test measure what it is supposed to measure?
  - ◆ For example:
    - The experimenter's 7th grade science test positively correlates with other 7th grade science achievement test results.
    - AND
    - The experimenters science test correlates to a higher degree with other science tests than it does with tests of other academic areas.

36

Evaluating Tests

◆ Standardized administration guidelines

◆ Appropriate vocabulary

◆ Clarity of directions

◆ Objectivity of scoring

37

Selecting & Administering a Test

◆ Standardized Measures
  - Library Resources
    - Mental Measurement Yearbooks Online
      - http://library.csus.edu/
    - Tests in Print
  - Electronic Resources ($15 per review)
    - Buros Institute
      - http://buros.org/mental-measurements-yearbook

38

Selecting & Administering a Test

◆ Make arrangements for testing in advance

◆ Create best possible test environment

◆ Be well prepared

◆ Protect test security

39

Portfolio Activity #4

Identify at least three (3) standardized measures relevant to areas of research interest. The following information should be included for each measure: (a) the name, publisher, and cost of the measure; (b) a brief description of what the measure purports to measure, (c) a brief summary of the measure's reliability and validity data.

Small group discussions

40

Portfolio Activities #5: Mini-proposals 1

- ◆ Briefly describe a survey research project relevant to one of their your research topics.
- ◆ Briefly describe a correlational research project relevant to one of your identified research topics

41

Next Week

- ◆ No Class (NASP
  - Complete CITI Human Subjects Research Course
- ◆ March 5
  - Descriptive Research
  - Read *Educational Research* Chapter 8
  - Portfolio Element #5 Due: Mini-proposal 1

42