Gathering Research Data

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

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

Scales of Measurement

<table>
<thead>
<tr>
<th>Scale</th>
<th>Properties</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Nominal        | Qualitative categories only. Observations sorted into categories by principle of equivalence. Scale categories differ from one another only in a qualitative sense. | Eye color
                |                                                                           | Gender
                |                                                                           | Ethnicity
                |                                                                           | Type of school
                |                                                                           | ADHD + no ADHD |
| Ordinal        | Observations are ranked in order of magnitude. Numerical ranks express a “greater than” relationship, but with no implication about how much greater. | Ordinal
                |                                                                           | 1 = Tallest 6’7”
                |                                                                           | 2 = 6’
                |                                                                           | 3 = 5’11”
                |                                                                           | 3 = 5’11”
                |                                                                           | 5 = 5’8”
                |                                                                           | 6 = Smallest 5’
Scales of Measurement

<table>
<thead>
<tr>
<th>Scale</th>
<th>Properties</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval</td>
<td>Numerical value assigned indicates order of merit and meaningfully reflects relative distances between points along the scale.</td>
<td>Educational Tests</td>
</tr>
<tr>
<td>Quantitative Variables</td>
<td>A given interval between measures has the same meaning at any point in the scale.</td>
<td></td>
</tr>
<tr>
<td>Ratio</td>
<td>Scale has all properties of an interval scale, and has an absolute zero point.</td>
<td>Length</td>
</tr>
<tr>
<td>Quantitative Variables</td>
<td></td>
<td>Weight</td>
</tr>
</tbody>
</table>

Scales of Measurement

- Family Income and Student Reading Test Scores
  - How is one quantitative (ratio) variable related to another quantitative (interval) variable?  
    - Correlation study

- Gender and Student Reading Test Scores
  - How is one Categorical (nominal) variable related to another quantitative (interval) variable?  
    - Correlation study

- Homework vs. Longer Classes and Math test Scores
  - How does one Categorical (nominal) variable affect another quantitative (interval) variable?  
    - Ex-Post Facto or Experimental Study

- ADHD (Y/N) and Reading comprehension test score
  - How is one Categorical (nominal) variable affect another quantitative (interval) variable?  
    - Ex-Post Facto study
Activity

- State a research question
- Identify the scale of measurement used in addressing the research question.
- Identify the relationship between variables.
- Identify the type of study

Group Comparison Variables

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

- Dependent Variable (the effect):
  - AKA Dependent Measure
  - The variable used to measure or assess the hypothesized effect.

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

- Homework vs. Longer Classes and Math Test Scores
  - How does one Categorical (nominal) variable affect another quantitative (interval) variable?

- ADHD and Reading comprehension
  - How does one Categorical (nominal) variable affect another quantitative (interval) variable?
Methods of Data Collection

- Standardized measures
  - e.g., published tests
  - These will be in your portfolio

- Experimental measures
  - i.e., measures developed by the researcher.

- Coding
  - of observations and records.

Types of Measuring Instruments

- Cognitive Tests (what people know and how they think)
  - Achievement Tests
  - Aptitude Tests

- Affective Tests (what people believe, feel, and perceive)
  - Attitude Scales
  - Interest Inventories
  - Personality Inventories
Methods of Data Collection

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

- **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)

**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.

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.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Art</td>
<td>✔</td>
</tr>
<tr>
<td>Transition</td>
<td>✔</td>
</tr>
<tr>
<td>Math</td>
<td>✔ ✔ ✔</td>
</tr>
<tr>
<td>L.A.</td>
<td>✔ ✔ ✔ ✔</td>
</tr>
<tr>
<td>Reading</td>
<td>✔ ✔ ✔ ✔ ✔ ✔</td>
</tr>
<tr>
<td>Free time</td>
<td></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Time</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00-8:15</td>
<td>✔</td>
</tr>
<tr>
<td>8:15-8:30</td>
<td>✔ ✔</td>
</tr>
<tr>
<td>8:30-8:45</td>
<td>✔ ✔ ✔</td>
</tr>
<tr>
<td>8:45-9:00</td>
<td>✔ ✔ ✔ ✔</td>
</tr>
<tr>
<td>9:00-9:15</td>
<td>✔ ✔ ✔ ✔ ✔ ✔</td>
</tr>
<tr>
<td>9:15-9:30</td>
<td></td>
</tr>
</tbody>
</table>

Systematic Observation: Data Collection

Event Frequency Data

Behavior event to be counted

<table>
<thead>
<tr>
<th>Date</th>
<th>Frequency</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.


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).

Systematic Observation: Data Collection

**Permanent Product Data**

- Behavioral outcome (or product) to be counted

<table>
<thead>
<tr>
<th>Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

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 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).


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.

### Systematic Observation: Data Collection

#### Interval Data (whole, partial, momentary)

<table>
<thead>
<tr>
<th>Time</th>
<th>Event 1</th>
<th>Event 2</th>
<th>Event 3</th>
<th>Event 4</th>
<th>Event 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 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.

### Evaluating Tests (Rudner, 1994)

- **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.
Evaluating Tests: Reliability

- The degree to which a test consistently measures whatever it purports to measure.
  - 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.

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.

Standard Error of Measure

- 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 takes SEM into account.
Evaluating Tests: Validity

- The degree to which a test measures what it is supposed to measure and, consequently, permits appropriate interpretations.
- Valid for specific purposes and populations.
- A matter of degree.

Evaluating Tests: Validity

- Content Validity
  - Degree to which 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.
    - e.g., the content validity of a science test would be determined by a group of experienced science teachers.

Evaluating Tests: Validity

- Criterion-Related Validity
  - The degree to which a test (the predictor) correlates with a second measure (the criterion).
    - **Concurrent Validity.** Both measures are administered in the same time frame. How well the measure reflects current functioning.
      - e.g., the correlation between the experimenter’s 7th science test results and student grades given by their 7th grade science teacher.
    - **Predictive Validity.** Both tests are administered at different times. How well the measure predicts future performance.
      - e.g., the correlation between the experimenter’s 7th grade science test results and student grades given by their 8th grade science teacher.
Evaluating Tests

Construct Validity

- The extent to which the test reflects the construct it is intended to measure. It requires a series of studies (including studies to determine content and criterion-related validity research), and is the most important form of validity.
- Does the test measure what it is supposed to measure.
  - e.g., the experimenter’s 7th grade science test positively correlates with other 7th grade science achievement test results. In addition, the experimenter’s science test correlates to a higher degree with other science tests than it does with tests of other academic areas.

Evaluating Tests

- Standardized administration guidelines
- Appropriate vocabulary
- Clarity of directions
- Objectivity of scoring

Selecting & Administering a Test

- Standardized Measures
  - Library Resources
    - Mental Measurement Yearbooks
      - http://library.csus.edu/
    - Tests in Print
  - Electronic Resources
    - Buros Institute
      - http://buros.unl.edu/buros/sp/search.jsp
    - Practical Assessment, Research & Evaluation
      - www.pareonline.net
Selecting & Administering a Test

- Make arrangements for testing in advance.
- Create best possible test environment.
- Be well prepared.
- Protect test security.

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
Portfolio Activities #5 & #6: Mini-proposals 1 & 2

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

Next Week

- Descriptive Research
- Read Educational Research Chapters 7 & 8.
- Portfolio Elements Due: (5) Mini-proposal 1.
- (6) Mini-proposal 2.