SAN JOSÉ STATE UNIVERSITY   **College of Social Work**
                                                                 S. W. 242
                                                              Spring 2008
                                                           Edward Cohen

Week 4: February 15, 2008
Design of Quantitative Survey Instruments / Introduction to SPSS and Descriptive Statistics

"There are three kinds of lies: lies, damned lies, and statistics."
--attributed to Benjamin Disreali (1804-1881)

Concepts You Should Know:
- Frequency (of a variable)
- Frequency distribution
- Absolute frequency distribution
- Cumulative frequency distribution
- Cumulative percentage frequency distribution
- Bar graph
- Histogram
- Frequency polygon
- Measures of central tendency
- Mode
- Median
- Mean (average)
- Measures of variability
- Range
- Percentile
- Interquartile range
- Variance
- Standard deviation
- SPSS Data Editor
- SPSS data view screen
- SPSS variable view screen
    - Variable label
    - Variable values
- "Click and point" method of data entry
- Syntax method of data entry
- SPSS output

I.   Design of Data Collection Instruments

   A. Review: asking the right questions:

         1. Avoid leading questions

2. Avoid double barreled questions

3. Match the language with age and cognitive levels of target population

4. Match the question with the intended cultural meaning of the target population

5. Build in ways to minimize error in response due to memory limitations

6. Remember—the attributes for "check one" questions should be *exhaustive and mutually exclusive*

7. Consider the order of questions in the instrument/survey:

   a)   Least sensitive (such as basic demographic info) to most sensitive

   b)   Whenever possible, questions should flow like a conversation

   c)   Prepare a few prompts for each question (could be in the instructions to the interviewer) to encourage response

B. A few important technical formatting principles:

1. See Rubin & Babbie, Chapter 9 for good examples of formatting

2. Number and identify sub-questions (1, 1a, 2, 2a, 2b etc.) to help respondents and research staff.

3. Allow sufficient space for responses and to maintain a "clean" look of the form. Keep general appearance in mind. Use large and clear font-types. Use colored paper covers (actual project version only). Leave space for data notes.

4. A vertical format is recommended with consistent patterns for codes, alignment, and use of response labels.

5. Include instructions and directions for question direction (e.g., "if yes go to…"). Assume that you will not be the only interviewer and that instructions are needed to maintain consistency of data collection.

6. Keep the set of response categories for a particular question all on same page

7. Make check boxes or lines that are clearly visible to an interviewer, a respondent, or someone doing data entry

8. Pilot test the instrument with a classmate or friend to make sure the questions make sense, are easily answerable, and flow in the intended order

C. Use of Existing Instruments—Aligning Your Variables

- In selecting the questions to ask about demographic characteristics and key variables, researchers should start with items comparable to those found in major databases or

published works. For example, in health surveys, variables are often measured according to the National Center for Health Statistics – National Health Interview Survey or related federally sponsored health surveys. (See Rubin & Babbie for examples.)

- Give examples about how the following variables can be assessed:

    - Household composition   - Age
    - Sex   - Marital Status
    - Race and Ethnicity   - Education
    - Employment Status   - Occupation
    - Socioeconomic Status   - Income

II.   Assignment # 3—Demographic Questionnaire

  A. Create an instrument to record data

  B. Can be used to record data for an interview you would be conducting, or

  C. Can be used for self-report by respondent. For this option, include appropriate instructions to the respondent

    1. Instructions on which attributes to check e.g. "Check only one" or "Check all that apply"

    2. Instructions to circle number or check box, when applicable

    3. Script to explain question and choices, as applicable (e.g. "Listed below are categories of ethnicity. Check the box of the ethnic category you most identify with.")

III.   Statistics I

  A. Counting

    1. **Frequency (of a variable)** – a count of observations falling in a value category (or in an attribute) of a variable, e.g. "the sample consists of 40 men and 60 women"

        a)   **Absolute frequency distribution**—simple counts

        b)   **Cumulative frequency distribution—**adds a column showing a running summed count

        c)   **Percentage frequency distribution—**adds a column showing the percentage of total observations in the sample. Another column could be percentage cumulative frequency distribution.

    2. Frequencies can be displayed graphically, with **frequency distributions** – tables or graphs the present the number (frequency) with which different values

(attributes) of a variable occur

    a)   **Bar graph** – represents nominal-level data (usually) with height (or length) of bars representing counts

    b)   **Frequency polygon** – graphical representation using dots and connecting lines to display the shape of the distribution of a ratio or interval level data

    c)   **Histogram** – like a bar graph, except the bars touch. Used to display frequency distributions of ratio, interval or ordinal data

3. **Interquartile range** – the number of observations (values) between the 25th and 75th **percentile** (-- which indicates the number of cases that fall below a certain value)

B. What's "typical"? Measures of central tendency

Measures of central tendency – we want to be able to describe what is "typical." In statistics, we describe typical through measures of central tendency (mode, median, mean). Why?
- They summarize data; one number explains a lot
- They provide a common reference point for comparing two groups of data

1. **Mode** – the value in a distribution of values within a data set that occurs most frequently (also see book examples)

0 0 0 1 1 2 2 3 3 3 3 3 4 4 5   N =   Mode =

0 0 0 0 0 1 2 2 3 4 5 5 6 7 7 7 7 7 8 9 11 14   N =   Mode =

2. **Median** – if data can be ranked into a list or array (interval or ratio variables) the median is that value which divides that list or array of values into two equal halves (also see book examples)

1 2 4 5 6 9 9   N =   Median =

1 3 4 4 4 5 5 5 6 7   N =   Median =

True of False: *The Median is the same thing as the 50th percentile*

3. **Mean** – called the arithmetic mean or average, is the sum of the values (interval or ratio variables) divided by the total number of values (also see book examples)

6 6 8 8   N =
                                                Mean =

The mean is sensitive to **outliers** – very extreme values in a frequency distribution

6 6 8 8 32   N =
                                                Mean =

    C. How much variation? Measures of variability (or dispersion)

        1. **Range** – (the largest value minus the smallest value + 1)

        2. **Variance** – the average value of the squared deviations from the mean

        3. **Standard deviation** -- the square root of the variance

IV.   Introduction to SPSS--See handout "Lab: SPSS Basics and Descriptive Statistics")

Class Process Questionnaire -- ScWk 242 Section: __3
__4

Date of Class: _____

Instructions—near the end of class or at anytime after class please complete *any* of these questions and return to the Instructor. **This is *not required* for grade or class participation.** Do not put your name on this sheet.

1. At what point in this class did you feel most engaged?

2. At what point in this class did feel the least engaged?

3. What action or discussion by anyone in the room did you find most helpful?

4. What action or discussion by anyone in the room did you find most confusing?

5. What surprised you most about class today?

6. Is there a topic that we should definitely get back to in the next weeks? Or, any particular types of teaching techniques you would recommend?