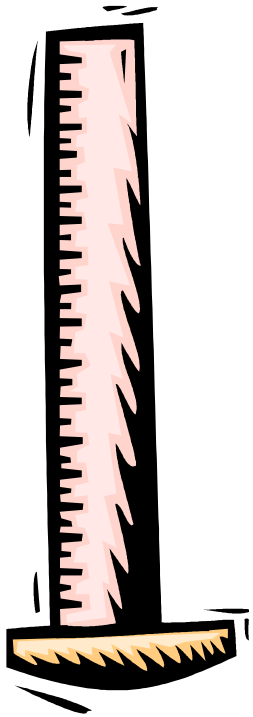
Evidence-Based Research & Sampling

Michael C. Rodriguez

The College of Education & Human Development

University of Minnesota

Reducing Error



- Measurement Error
- Nonresponse Error
- Coverage Error
- Sampling Error

Measurement Error

- The result of asking poorly worded questions so that responses are inaccurate or uninterpretable

  - Example:
    - The researcher asked participants about "*how the social interaction level had changed in the past year*" using these choices:

      (1) increased a lot
      (2) increased somewhat
      (3) increased a little
      (4) decreased

Nonresponse Error

- The result of people who do not respond to a survey being different from those who do respond -- different in a way that is meaningful to the study

  - Example:
    - 80% of the residents that were college graduates responded while only 20% of residents high school graduates only responded -- particularly important if a goal of the researcher was to compare respondents based on level of education

Coverage Error

- The result of not giving every member of the population a chance to be included in the sample

  - Example:
    - The researcher decided to survey hundreds of residents in each neighborhood using the telephone directory – not realizing that some neighborhoods have large numbers of households without phones or unlisted numbers.

Sampling Error

- The result of only including some of the population in the study -- literally because we only have a sample

- Example:
  - The researcher decided to survey 240 employers in her state about their satisfaction with the work readiness of recent high school graduates.

Reducing Nonresponse Error

- Respondent friendly instruments, procedures
- Multiple contacts
  - Prenotice letter
  - Instrument
  - Postcard thank you/reminder
  - Replacement instruments
  - Special mail, phone contacts
- Return envelope with real stamps
- Personalized correspondence
- Token prepaid incentives

Reducing Coverage Error

- Elements of a Sampling Frame
  - Contains the entire target population
  - Eliminates individuals not in the target population
  - Maintained & updated
  - No repeated names or addresses
  - Contains other information that may contribute to the research effort

Reducing Sampling Error

- Two conditions when trying to estimate a population characteristic
  - The smaller the sample, the less precise the estimate
  - The more variable the sample (population), the less precise the estimate

Probability Sampling Methods

- Simple Random Sampling
- Stratified Sampling
- Cluster Sampling
- Complex Multistage Sampling

Simple Random Sampling

- Each individual has an equal opportunity to be selected in the sample

- Why?
  - To allow us to generalize
  - So we can apply the laws of probability (statistics)

Simple Random Sampling

- We can estimate population values by computing simple means, standard deviations, percentages or proportions, and a variety of other statistics (correlations, R^2s, mean differences, etc.)

  AND

- We can compute the precision of our estimates by relying on estimates of sampling error – called standard errors in statistics, and then report confidence intervals, which are more accurate portrayals of our findings

Stratified Random Sampling

- Subdivide the population into subgroups or strata
  - Choose strata for planned comparisons
- Sample within each strata

- Why?
  - Protection from bad samples
  - Achieve precision for subgroups
  - May be more convenient with lower cost
  - Provide greater precision for all estimates

Stratified Random Sampling

- What is the result?
  - Observations within strata tend to be more alike than in the population as a whole; reduction in variance within strata reduces the variance for the population estimates – resulting in greater precision

- Population total then becomes the weighted average of sample stratum estimates (by relative strata size) variances are based on the sum of strata variances

Stratified Random Sampling

- Allocating Observations to Strata
  - Proportional
    - Same proportion from each strata
  - Optimal
    - Most information at the least cost
    - Achieve a certain level of precision per strata
- Larger samples within strata are called for when
  - The stratum accounts for a large part of the population
  - The variability within a stratum is high
  - Sampling within the stratum is inexpensive

Cluster Sampling

- Randomly select groups or clusters
  - Natural or predefined clusters
  - Addresses administrative barriers and ethics
- May or may not sample within clusters

- Why?
  - Constructing a sampling frame is difficult, expensive, or impossible; population is spread out geographically or may naturally exist in clusters

Cluster Sampling

- What is the result?
  - Cluster sampling tends to decrease precision because we obtain less information per observation – individuals within clusters are more likely to be similar than those across clusters.
  - When we sample everyone in a cluster, we are partially obtaining repetitive information instead of new information – resulting in less precise estimates of population characteristics.

Cluster Sampling

- One-Stage Cluster Sampling
  - Equal size clusters
    - We can consider clusters as observations and use SRS methods
  - Unequal size clusters
    - We can weight cluster characteristics based on the proportion of the population represented by each cluster

We need to know cluster size and the total population size

Cluster Sampling

- Two-Stage Cluster Sampling
  - We randomly sample within clusters
  - This results in sampling error at two stages
    - Sampling clusters
    - Sampling within clusters

Nonprobability Sampling

- Systematic sampling
- Convenience sampling
- Snowball sampling
- Purposive sampling
- Quota sampling
- Focus groups
- Expert panels

Sample Size

- What is the null hypothesis?
- What is the desired level of significance related to the null hypothesis involving the mean in the population?
- What chance should there be of detecting an actual difference? Power.
- What differences between means are important?
- What is a good estimate of the standard deviation in the population?

One final source of error that is impossible to account for, adjust for, or control, because it is entirely random and not predictable, but is always present in social science research is …

free will

Dillman, D. A. (2000). *Mail and internet surveys* (2nd ed.). New York: Wiley & Sons, Inc.

Lohr, S. (2001). *Sampling Design and Analysis*.

Schonlau, M. Fricker, R. D., Jr., Elliott, M. N. (2002). *Conducting research surveys via e-mail and the web*. Santa Monica, CA: RAND.

http://measurement.class.umn.edu/survey/nrccte.html