Harvard University
Program on Survey Research

How to Frame and Explain the Survey Data in your Honors Thesis

Chase H. Harrison Ph.D.
Program on Survey Research
Harvard University

Surveys

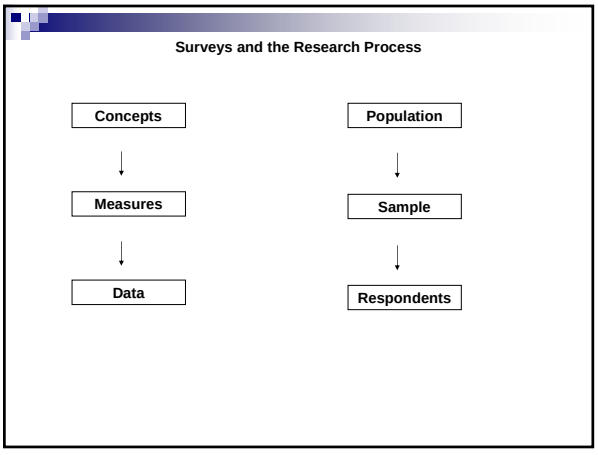
- Systematic method of data collection
- Usually use samples
- Designed to measure things
  - Attitudes
  - Behaviors
- Create statistics
  - Descriptive
  - Analytic

Overview of Research Process

Research Theories → **Survey Methods** → **Reporting And Analysis**

Types of Surveys

- Original survey you designed yourself
- Non-distributed private survey
- Archived survey
- Survey data
  - Paper (or electronic) report
  - Questions from database

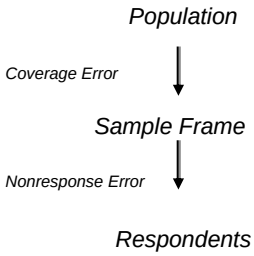


Samples and Populations

Survey Sampling

- *A **census** attempts to collect data from all members of a population.*
- ***Random samples** let you use collect data from a portion of a population and use sampling statistics to generalize your findings to a large population.*

Survey Sampling



Probability Samples

- ***Based on Probability Theory***
- ***Allow Inference to Sample Frame***
- ***Sample Variance and Error Can Be Calculated***
  - Sample Records Are Drawn From a Well-Specified Frame
  - Sample Records Are Drawn According to Well-Specified Procedures With Known Properties
  - Each Sample Record Has a Known Non-Zero Probability of Selection
  - Data are Adjusted (Weighted) As Required To Reflect Sample Design

Non-Probability Samples

- Availability Samples
  - Convenience Samples
  - Volunteer Cases
- Purposive Cases
  - Typical Cases
  - Critical Cases
  - Snowball Samples
- Quota Samples

Sample Frames

- List or a set of procedures
- Sometimes requires two or more stages of selection
- Designed to cover target population

Stratified Samples

- Divide sample records into similar groups
  - Proportionate Stratification represents each stratum in proportion to its prevalence in the population
  - Disproportionate Stratification samples groups with non-proportionate probabilities
    - Some groups are oversampled
    - Stratification might need to be adjusted or weighted for total population estimates

Sampling Procedures

- Multi-stage cluster samples
  - Select a Primary Sampling Unit (PSU) and then conduct further sampling
- Systematic samples (sample every n'th person in the frame)
- Simple random samples or Equal Probability Selection Method (EPSEM) samples give each sample record an equal probability of being selected.

Samples and Units

- Does sample record correspond to population unit?
  - Household versus person
  - Telephone household versus household
  - Organization versus employee
- How are reporting units selected
  - All interviewed
  - Random selection
  - Convenience selection
- Nature of Information
  - Sometimes information is collected from administrative records
  - Sometimes, multiple respondents are needed to answer questionnaires
  - Sometimes, proxy respondents are used

Sample Error

- Based on Statistical Theory
- Describes Variability
- Applies From Respondents to Sample Frame

Coverage Error

- People excluded from sample frame
- Typical sampling statistics assume no coverage error
- Bias:
  - Proportion Excluded
  - Differences Between Excluded and Included

Nonresponse Error

- Sample Members Who Do Not Respond
  - Reasons:
    - *Unable*
    - *Unavailable*
    - *Unwilling*
  - Bias:
    - Proportion Excluded
    - Differences Between Excluded and Included

Response and Nonresponse

- Percentage of Valid Sample Records that Are Included in Statistic
  - Unit Nonresponse = Missing Respondents
  - Item Nonresponse = Missing Answers

Evaluating Coverage and Nonresponse Bias

- Evaluate magnitude of exclusion
  - Percent of population excluded from sample frame
  - Percent of sample frame non-responding
- Evaluate or discuss potential differences on key variables
  - Measurement of survey variables non-covered/responders is difficult
  - Compare with population or sample frame statistics if known
  - Adjusting or weighting data is possible
  - Reasonably discuss potential differences if exclusion is large

Measuring Levels of Nonresponse

- Response Rates
  - The percentage of eligible members of your sample who completed your survey
- Coöperation Rate
  - The percentage of (eligible) people you contacted who participated in your survey.

Outcome Rate Standards

- American Association for Public Opinion Research (AAPOR)
  - Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 4th Edition (Kenexa, KS: Feb. 2006)
  - Different Specific Rates for:
    - RDD Telephone Surveys
    - In-Person Household Surveys
    - Mail Surveys of Specifically Named Persons
    - Internet Surveys of Specifically Named Persons
  - Guidelines for Similar Surveys
- http://www.aapor.org/pdfs/standarddefs_4.pdf

Field Procedures

Modes

- Interviewer Administered Questionnaire
  - Face-to-face
  - Telephone
- Self-Administered Questionnaire
  - Mail
  - Web (or other computer)
  - Intercept (describe role of interviewer)
- Multiple Modes
  - For same respondent
  - For different respondents

Field Protocols

- What rules or procedures were used to collect data?
- How were respondents contacted?
- Who contacted the respondents (if by mail or telephone)
- When were respondents contacted? (Time period of survey)
- What happened when sampled units were unavailable or refused?
- How many times were respondents contacted?

Field Protocols

- What instructions were given to interviewers? (if used)
- What instructions were given to respondents?
- Protocol Clarification
  - Respondent questions
  - Interviewer questions
- What incentives or inducements were used?

Questions and Measures

Theories and Surveys

- Concepts (Theoretical Ideas)
- Measures (Questions or Scales)
- Statistics (i.e. Data)

Concepts

- **Broad Theories**
- **Meaningful**
- **Rich**

Measures

- **Specifically Operationalized**
- **Bounded by**
  - **Content**
  - **Scope**

Types of Measures

- Factual
  - Behavior
  - Dates and Duration
  - Demographic
- Attitudes
  - Values
  - Judgments
  - Opinions

Measuring Attitudes

- Latent Construct
- Can consist of several facets or aspects
- Questions are often scaled
- Scales can be created from multiple batteries of questions

Reliability and Validity

- Reliability
  - The ability of a question to produce consistent results over repeated trials
    - Different times
    - Different surveys
- Validity
  - The ability of a measure to accurately measure what it is trying to measure
  - Construct validity measures the extent that a question measures the underlying construct it is intended to measure

Types of Questions

- Open End
- Closed End
- Discrete (yes/no)
- Rating Scale

Types of Measures

- Interval / Continuous
  - Every possible value included
- Ordinal
  - All values can be placed above or below one another
- Nominal
  - Unique discrete categories

Questions

- Should your percentages include or exclude people who say “don’t know” from the base?
- Should your percentages include or exclude people who didn’t answer the question from the base?

Response Effects

- Social desirability
  - Tendency varies across cultures
  - Topic Sensitivity varies across cultures
- Acquiescence
  - Tendency to always say “yes”
- Use of scale extremes
  - Giving extremely high or low answers
- Use of “no opinion” options

Response Effects

- Primacy effect
  - Respondents focusing on initial items or response choices
  - Typical in self-administered surveys
- Recency Effect
  - Respondents focusing on most recent thing they heard
  - More common in interviewer-administered surveys.

Analyzing Attitude Questions

- Percentage
  - One category
  - Two Collapsed Categories
- Numeric
  - “Mean number”
  - Realize this is an ordinal mean
  - Numeric scale

Creating Scales from Multiple Questions

- Possible to create scales from multiple questions
- Can measure activities or attitudes
- Often treated as interval data
  - Mean or Median can be reported
- Sometimes scaled to 1, 10, or 100

Documentation

- Best to discuss all decisions either in text or in appendix.
- Full question wording should be given, either in text or appendix.
- Additional documents:
  - Full questionnaire
  - Pre-notification and contact letters
  - Specialized interviewer instructions

Essential Elements

- Mode or method of data collection
- Dates and geography of data collection
- Description of target population
- Description of sample frame and sample methods
- Characteristics of respondents

Sample Documentation

- Universe study is intended to represent
- Description of sample frame and source
- Description of sample design:
  - Cluster size
  - Number of callbacks
  - Eligibility criteria and screening procedures
  - Other pertinent information

Respondents and Response Rates

- Size of samples and number of respondents
- Demographic profiles of respondents
- Response or completion rates
- Comparison of respondent characteristics to sample or population characteristics

Questionnaire Elements

- Methods for developing questionnaire
- Sources of questions if appropriate
- Full wording of all questions
  - Include visual exhibits
  - Include preceding instructions
  - Include explanation to the interviewer or respondents
- Description of data adjustment or indexing
- Description of coding methods and categories if appropriate

Resources at Harvard

- *General Resources:*
  http://www.ig.harvard.edu/psr/internet_resources

- *Specific Resources at Harvard:*
  http://www.ig.harvard.edu/psr/psr_resources_tips

Harvard University
Program on Survey Research

How to Frame and Explain the Survey Data in your Honors Thesis

Questions and Discussion