SURVEY DESIGN AND ANALYSIS

Dominique Zéphyr
TODAY’S PRESENTATION

• Intro to Survey Research
  • When to use a survey?
  • Types of survey

• Issues in Survey Research
  • Select the sample
  • Question design and type
  • Non-response
  • Weighting
WHEN TO USE A SURVEY TYPES OF SURVEY
WHAT IS A SURVEY?

- Today the word "survey" is used most often to describe a method of gathering information from only a portion of a population of interest (i.e. Sample).

- Survey research is about asking questions about:
  - Behaviors
  - Opinions/Attitudes
  - Beliefs
  - Symptoms
  - Attributes (Demographic Characteristics)
WHAT A SINGLE SURVEY CAN TELL YOU

- Collects data at single point in time
- Can gather wide variety of data
- Collects data allowing calculation of prevalence or incidence rates
- Can be done when other data collection systems (e.g. clinical observation) not feasible
WHAT A SINGLE SURVEY CANNOT TELL YOU

- Cause-effect relationships
- Difficult to answer “why” questions
- Must be repeated to follow trends over time
WHEN TO USE SURVEYS

- Surveys cost money, manpower, and time

- Think about alternative sources
  - Information from non-survey sources
  - Information from existing survey data

- Carefully consider if survey necessary before beginning
STILL THINK YOU NEED TO DO A SURVEY?

○ Vital to define research questions or survey objectives
  • What exactly do I want to find out?
  • Is the hypothesis specific enough to be researchable and yet still meaningful?
  • What does the relevant literature in the field indicate about this problem?
  • Which are the independent and which are the dependent variables?
  • Are there control variables?
The Survey Process

- Identify the goals and objectives
- Develop the research design
- Identify the population and select the sample
- Design the questionnaire
- Pre-test the survey
- Administer the survey
- Clean the data
- Analyze and interpret the results
- Write up the results
- Present the finding
BASIC SURVEY FORMATS

- Self-administered (Mail, Internet and Email survey)
- Telephone interview
- Face-to-face interview
<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-completion</td>
<td>• Cheap</td>
<td>• Low response rate (and possible bias from this)</td>
</tr>
<tr>
<td></td>
<td>• Cover wide area</td>
<td>• Questions need to be simple</td>
</tr>
<tr>
<td></td>
<td>• Anonymity protected</td>
<td>• No control over interpretation</td>
</tr>
<tr>
<td></td>
<td>• Interviewer bias doesn’t interfere</td>
<td>• No control over who fills it in</td>
</tr>
<tr>
<td></td>
<td>• People can take their time</td>
<td>• Slow</td>
</tr>
<tr>
<td>Telephone survey</td>
<td>• Can do it all from one place</td>
<td>• People may not have home phones/be ex-directory</td>
</tr>
<tr>
<td></td>
<td>• Can clarify answers</td>
<td>• You may get wrong person or call at wrong time</td>
</tr>
<tr>
<td></td>
<td>• People may be relatively happy to talk on the phone</td>
<td>• May be a bias from whose name is listed/who’s at home</td>
</tr>
<tr>
<td></td>
<td>• Relatively cheap</td>
<td>• Easy for people to break off</td>
</tr>
<tr>
<td></td>
<td>• Quick</td>
<td>• No context to interview</td>
</tr>
<tr>
<td>Face-to-face interview</td>
<td>• High response rate</td>
<td>• Slow</td>
</tr>
<tr>
<td></td>
<td>• High control of the interview situation</td>
<td>• Expensive</td>
</tr>
<tr>
<td></td>
<td>• Ability to clarify responses</td>
<td>• Interviewer presence may influence way questions are answered</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If there is more than one interviewer, they may have different effects</td>
</tr>
</tbody>
</table>
ISSUES

- Select the sample
- Question Design and Type
- Non-Response
- Weighting
SELECT THE SAMPLE
Who do you need to survey?

- Everyone in a specific target population? (A census)
  
  If you
  - Have a relatively small population
  - Have reason to expect a relatively low response rate
  - Want to have enough respondents to do sub-group analyses for small groups

- A sample of the target population

  If the population is
  - Too big and to do so would be too costly
  - Too time consuming, or impossible.
HOW MANY PEOPLE TO SURVEY?

- **Sample size?** Depends on:
  - Cost
  - Anticipated response rate
  - Number of responses needed per subgroup interested in
  - Level of precision you want ("margin of error")

- **Create a “sampling frame”**
  A list of elements from which the sample will be selected (i.e. members of the population)
Sampling

- Probability sampling: Every element in the population has a known non-zero chance of selection

- Such a sample allows us to make inferences about the population as a whole, based on the sample results

- Different from “convenience sampling” (e.g., snowball sampling)
PROBABILITY SAMPLING METHODS

- **Simple random sample**
  - List elements in the sampling frame and use random number chart or computer program to select elements

- **Systematic sampling**
  - List elements in the sampling frame and take every Nth element, based on how many needed in sample

- **Stratified sampling**
  - Draw appropriate proportion of elements from homogenous groups interested in
**BAD SAMPLE = BAD DATA**

- Survey results are only representative of the sampling frame from which the sample was selected.

- Sample survey results are generalizable only if respondents are randomly selected.

- Example: If you cannot control who accesses and responds to a web survey, the results are NOT generalizable to the population.
QUESTION DESIGN AND TYPE
TYPES OF QUESTIONS

- Open-ended
- Closed-ended
- Rating/Likert Scale
Open-End Questions

- Pros
  - Stimulate free thought
  - Solicit suggestions
  - Clarify positions
  - Richer information

- Cons
  - Respondent burden
  - Incomplete, irrelevant, uninterpretable responses
  - Coding, analyzing, reporting

- Limit the number of open-end questions BUT always include at least one for additional comments.
CLOSED-ENDED QUESTIONS

- Pros
  - Less demanding on respondent
  - Easy to analyze and make comparisons

- Cons
  - Limits richness of data
LIKERT SCALE

• Example:
• STATEMENT: This is a useful lecture
  – Agree strongly 1
  – Agree 2
  – Neither agree or disagree 3
  – Disagree 4
  – Strongly Disagree 5
QUESTIONS

- Good questions are connected to your survey objectives
  - What do you want to learn?
  - What pieces of specific information are needed to satisfy these objectives?

- Bad Question = Bad Data
  - The way a question is worded and the response options offered determine the nature of the data received.
Good Questions

- Good questions **reduce error**
  - By increasing the respondent’s willingness to answer
  - Increases reliability and validity.
    - Less error = better data
USE SIMPLE SENTENCE STRUCTURES

- DON’T: In the past 30 days, when you used needles for injecting drugs, have you shared them with anyone else?

- DO: In the past 30 days, have you shared needles?
**Avoid “Double-Barreled” Questions**

- **DON’T**: Are you getting along better with your family and friends?
- **DO**: Are you getting along better with your family? And Are you getting along better with your friends?
DO NOT USE DOUBLE NEGATIVES

DON’T: Do you oppose not allowing gays and lesbians to legally marry?

DO: Do you favor allowing gays and lesbians to legally marry?
**Good Questions…**

- Are clear and use simple language
- Are concise
- Are specific
- Are possible to answer
- Don’t overly tax the respondent’s memory
- Are not overly sensitive
- Are relevant to the respondent
- Do not use double negatives
- Avoid biased terms
THE QUESTIONNAIRE

- **Question order**
  - Start with interesting, easy, non-threatening questions

- **Length of questions**
  - Ensure questions are not too long and do not contain multiple concepts.

- **Length of questionnaire**
  - If long, you may not get meaningful data for questions at the end.

- **Appearance of questionnaire**
NON-RESPONSE
SURVEY NON-RESPONSE

- Non-response
  - Potential respondent can not or will not participate in the survey or answer specific questions

- Bias
  - Nonresponse bias is the bias that results when respondents differ in meaningful ways from nonrespondents.

- Strategies for Non-response
  - Initial contact – ‘warm’ approach
  - Short and simple questionnaire
  - Follow-up
  - Incentives
HOW TO HANDLE NON-RESPONSE ERRORS

- Generalize to the respondents only. In stating that results are accurate for those surveyed, you can avoid making incorrect inferences about the larger population.

- Assume there is no response bias and generalize to the population. If you know the population well and perceive the results to be reasonable, this strategy may be reasonable.

- Compare data in hand on respondents and nonrespondents. If data, e.g., sex, age, race, is available, the composition of respondents can be compared with that of nonrespondents to see if there are any differences.
WEIGHTING
WEIGHTING

- Weighting is used to compensate for

  - Unequal probabilities of selection -- Over-sampling of specific cases or disproportionate stratification

  - Nonresponse (typically, a unit that fails to respond) -- Propensity to respond may depend on age, race/ethnicity, gender, place of residence

  - In post-stratification to adjust weighted sample distributions for certain variables to make them conform to the known population distribution
WEIGHTS IN SAS

- Use the weight statement with the normal procedures
  - Examples:
    ```sas
    proc means ;
    var bpdia;
    weight fwgtexam ;
    run ;
    ```

- Use specific statistical procedures that analyze data from complex sample surveys
  - `SURVEYFREQ`, `SURVEYLOGISTIC`, `SURVEYMEANS`, `SURVEYREG`, and `SURVEYSELECT` procedures
    ```sas
    proc surveyfreq data=one;
    tables  Response;
    strata State ;
    cluster School;
    weight SamplingWeight;
    run;
    ```