Survey Design: Theory & Best Practices

UWGH Coffee & Quality

Jie Wu
August 6, 2020
www.datahouston.org
In the past week, how many virtual meetings have you participated in?
Before Survey Design

• What is the purpose of the survey? What is your research question? What are the topics?
• Is a survey the best way to collect the information that you need?
• What do you plan to do with the survey data? => An Analysis Plan
Before Survey Design

• Who is your target population? Are there any particular subgroups?
  • Understand your respondents and show respect
  • How to motivate your respondents to participate and give you the best information?
Example

COVID-19 Response Parent Survey

- Purpose: The survey is to gather information and inform decision-making for next academic year.
- Target population: Parents
Exercise #1

In the past week, how many virtual meetings have you participated in?

- Why do you ask this question?
- What are you trying to measure?
- Is this the best way to measure?
- How will you analyze the data?
Exercise #1

In the past week, how many virtual meetings have you participated in?

◉ If you were a respondent, what are the possible interpretations of this question?
◉ Is there a socially desirable response?
◉ What’s the context of the question?
Questions
Questions

• Questions that are easy for your respondents to provide accurate and reliable answers
• Questions that can be asked and answered quickly
• Less burden on respondents
Exercise #2.1

Was your home damaged at all as a result of the hurricane?

- How to improve the question to avoid misinterpretation?
Exercise #2.2

Was that minor damage or major damage?

- How to improve the question to avoid misinterpretation?
Exercise #3.1 (5 minutes)

Can you come up with three questions that you could ask an individual to determine whether or not he will vote in November?
Answers

• How difficult it is to provide an answer?
• What is your respondent’s ability? E.g. familiarity with the topic
• Do your respondents have motivation to provide an answer? E.g. belief about the importance of the survey
Response Options

• Order of response options is important
• Don’t Know
Question Types and Response Options

• Closed-ended questions
  • Dichotomous (yes/no, agree/disagree, for/against)
  • Ratings (Likert scale)
  • List of items (multiple choice)
  • Ranking/Ordinal
  • Skip logic and contingency questions
Question Types and Response Options

• Open-ended questions
  • Pros: Responses to open-ended questions are often more valid and reliable.
  • Cons: More time to record answers and code the responses
• Semi closed-ended questions
Exercise #3.2 (10 minutes)

Please write response options for each of your questions.

Discuss the pros and cons of your current design.

Do you want to change the wording of the questions?
How to avoid questionnaire problems?

- Pretest, Pretest, Pretest!
  - To identify questions causing confusion
  - Wording and grammar issues
  - Potential biases
  - Leading questions or presumptive questions
  - Uninformative questions
  - Typos, broken skip logic, issues with response categories
Question Wording

• Avoid double-barreled questions
  • E.g. “How well do you get along with your supervisors and coworkers?” “Is this workshop informative and interactive?”

• Avoid leading questions
  • E.g. “What size do you want?” vs. “Do you want Grande?”
    • “Do you want to add a warranty to your car?” “Shall I sign you up for the 3-year warranty or the 5-year warranty?”

• Avoid negative and double negative questions
  • E.g. “I don’t know about this program. True or False.” vs. “How well do you know about this program?”
Question Wording

• Simple words and sentences that are easy to comprehend
• Clear definitions, no jargon
• Be specific and neutral
• Allow for all possible responses
Question Order

• Easy questions at the beginning
• Ask open-ended questions early
• Ask sensitive questions toward the end
• Group questions by topics vs. randomize the order
Criteria

• Reliability – whether the survey instrument produces consistent results
• Validity – whether the survey instrument measures what we intend to measure
Using an existing question vs. Creating a new question

• Use an established question with known reliability, validity, and the universe of response categories
• Check if this is the right question for your purpose and population.
Resources:

☞ Kinder Houston Area Survey
   https://www.datahouston.org/khas

☞ General Social Survey (GSS) since 1972: http://gss.norc.org/

☞ Census Bureau Questionnaire Archive:
   https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive.html
Resources:

☞ Pew Research Center since 1997
   https://www.pewresearch.org/question-search/

☞ CDC: https://www.cdc.gov/nchs/data_access/ftp_data.htm

☞ Kaiser Family Foundation Polling: https://www.kff.org/polling/

☞ National Household Education Surveys: https://nces.ed.gov/nhes/
Resources:

☞ Roper iPoll Archive (Roper Center for Public Opinion Research):
   https://ropercenter.cornell.edu/ipoll/

☞ Political Polls since 1972
   https://ropercenter.cornell.edu/data-highlights/elections-and-presidents

☞ Religious Data Archives: http://www.thearda.com/archive/browse.asp

☞ Rand American Life Panel study:
Building Better Neighborhoods

Survey

Sampling

Target Population

Sample
Sample Size

- Determine the number of participants in a sample group (sample size)
  - Estimated number of people in the target population
  - Representativeness
  - Budget
  - Time available
- E.g. \( N = \frac{t^2 \times P \times (1-P)}{m^2} \)
  - \( N \): required sample size
  - \( t \): confidence level at 95% (standard normal distribution: 1.96)
  - \( P \): estimated prevalence of the variable of interest (e.g. 50% of the customers are satisfied with the service)*
  - \( m \): margin of error (e.g. at 5%)
Example:

<table>
<thead>
<tr>
<th>Survey Sample Size (N)</th>
<th>Margin of Error (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,000</td>
<td>2%</td>
</tr>
<tr>
<td>1,000</td>
<td>3%</td>
</tr>
<tr>
<td>700</td>
<td>4%</td>
</tr>
<tr>
<td>500</td>
<td>4%</td>
</tr>
<tr>
<td>400</td>
<td>5%</td>
</tr>
<tr>
<td>300</td>
<td>6%</td>
</tr>
<tr>
<td>200</td>
<td>7%</td>
</tr>
<tr>
<td>100</td>
<td>10%</td>
</tr>
<tr>
<td>50</td>
<td>14%</td>
</tr>
</tbody>
</table>

Margin of error at a 95% level of confidence if the estimated prevalence of the variable of interest is 50%
Example

“xxx ISD Community,

Please receive this as a revision to the update and survey you received yesterday evening.

We regret to share that the settings on last night’s survey did not limit respondents from completing the survey multiple times. Unfortunately, it has become clear that this limitation has skewed and invalidated the current results.”
## Interviews

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
</table>
| **Personal Interviews – focus groups; stakeholder interviews (informants)** | • Face to face contact  
• More data and reliable information  
• Personal information can be obtained or verified easily | • Expensive and time consuming esp. when large and widely spread geographical sample is taken  
• Certain types of respondents may not be easily accessible  
• Respondent may give wrong information |
| **Telephone interviews** | • Faster, cheaper and less time consuming  
• Recorded  
• Computer-assisted telephone interviewing system  
• Wider representation of sample is possible | • Restricted to respondents who have landline or cellphone  
• Little time is given to respondents for answers  
• Not suitable for intensive surveys |
## Interviews

<table>
<thead>
<tr>
<th>Method</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
</table>
| Questionnaire                  | • Low cost even when the universe is large and widespread  
  • Free from bias of interviewer  
  • Respondents are given enough time to give well thought answers                                                                                                                                         | • Low response rate; incomplete  
  • It can be used only when the respondents are educated and cooperative  
  • Difficult to know if respondents are truly representative  
  • slow                                                                                                                                                                                                  |
| Schedule method – e.g. population census | • Useful in extensive enquiries  
  • Fairly reliable results                                                                                                                                                                              | • Very expensive  
  • Usually adopted in investigations conducted by government agencies                                                                                                                                 |

---

**Low cost** even when the universe is large and widespread.

**Free from bias of interviewer.**

**Respondents are given enough time to give well thought answers.**

**Low response rate; incomplete.**

**It can be used only when the respondents are educated and cooperative.**

**Difficult to know if respondents are truly representative.**

**Slow.**
Modes

• Mixed mode using both traditional mail and a web version vs. paper-only
  • Offering the web version may not increase the response rate, but rather shifting replies from the paper version to the web version.
  • Easy transfer of electronically collected data into statistical packages

• Increase the response rate:
  • Increase contact with the respondents
Sampling Methods

• Probability sampling
  • Simple random sampling (SRS)
  • Systematic sampling: e.g. every 10\textsuperscript{th} person on the list
  • Stratified random sampling: e.g. race
  • Cluster sampling: e.g. block
Example: KHAS

- **Stratified random sampling**
  - Interviews by Sample Type & Ethnicity
  - Margin of Error and Design Effect by Selected Groups

<table>
<thead>
<tr>
<th></th>
<th>Total Interviews</th>
<th>Landline</th>
<th>Cell Phone</th>
<th>Design Effect</th>
<th>Margin of Sampling Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>460</td>
<td>269</td>
<td>191</td>
<td>2.58</td>
<td>7.3</td>
</tr>
<tr>
<td>African American</td>
<td>217</td>
<td>119</td>
<td>98</td>
<td>2.23</td>
<td>9.9</td>
</tr>
<tr>
<td>Hispanic</td>
<td>215</td>
<td>73</td>
<td>142</td>
<td>2.10</td>
<td>9.7</td>
</tr>
<tr>
<td>Other</td>
<td>109</td>
<td>47</td>
<td>62</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>1001</td>
<td>508</td>
<td>493</td>
<td>2.61</td>
<td>5.0</td>
</tr>
</tbody>
</table>
Sampling Methods

- Non-probability sampling
  - Convenience sampling: e.g. intercept survey
  - Quota sampling
  - Snowball sampling
  - Judgement/Purposive sampling
  - Less costly, but do not support statistical interference
  - Results might be useful in developing research hypotheses or identifying issues
Bias

• Frame coverage bias
• Selection bias
• Size bias
• Nonresponse bias
Incentives, Consent, Confidentiality
How to select a survey vendor

- Cost
- Service: # of contacts, geocoding, statistician
- Training process
- Past projects
Best Practices for Survey Research

- Have specific goals.
- Consider alternative data beyond a survey.
- Select samples that well represent the population to be studied.
- Use designs that balance costs with errors.
- Take great care in matching question format and wording to the concepts being measured and the population being studied.
- Pretest questionnaires and procedures.
- Train interviewers carefully on interviewing techniques and the subject matter of the survey.
- Check quality at each stage.
- Maximize cooperation or response rates within the limits of ethical treatment of human subjects.
- Use appropriate statistical analytic and reporting techniques.
- Develop and fulfill pledges of confidentiality given to respondents.
- Disclose all methods of the survey to allow for evaluation and replication.

The Kinder Institute’s HCDC program seeks to facilitate the practical use of data by community leaders for decision-making and capacity building in Houston’s disadvantaged communities.

www.datahouston.org