Isha Ray
Energy & Resources Group; UC Berkeley
isharay@berkeley.edu

DESIGNING SURVEYS

Surveys

- One of the most important tools for data collection and measurement in social science research

Surveys are Q&A formats that involve:

- Large sample sizes (usually)
- Random samples (if aiming to represent the population)
- <u>Structured</u> questionnaires (set questions; pre-coded categories; set order of questions; no prompting or probing)

Together these help to:

1. reduce variability / error due to the interviewer;

(Useful because of e.g. poorly worded questions, memory problems, inaccurate or inconsistent recording or coding BUT prevent spontaneous follow-ups)

2. aggregate the responses & represent population of interest; and

3. enable quantitative analysis when the original data are qualitative

"Individuals vary, but percentages remain constant. So says the statistician"

[Sign of Four]

Surveys are not open-ended interviews

- You structure the questions -- you cannot deviate from respondent to respondent
- You cannot alter course in a burst of inspiration
- You use pre-coded categories
- You do not allow the respondent to speak in her "own" voice
- You use large sample sizes (ideally)
- You can now allow aggregate responses & run stats on even "qualitative" data so you can "represent" a bigger population -- of course your stats are only as representative as your sample

Selecting Survey Method (1)

- Selecting the type of survey is critical (mail; phone; internet; face-to-face)
- What guides your decisions?

Population Issues
- Can you enumerate the population?
- Is the population literate?
- Is it dispersed?

Sampling issues
- Can you find the respondents easily?
- What is your unit of analysis?
- How likely are respondents to drop off? (why important?)

Question issues
- What types of questions will you ask? (e.g. embarrassing ones?)
- Are the questions complex? Is the questionnaire long?
- Will you need filters?

Selecting Survey Method (2)

Content issues

- Can the respondents be expected to know about the issues that you are studying?
- Might the respondent need to consult records?

Bias issues

- Can you avoid respondents trying to please you?
- Can you be sure <u>who</u> is responding? (related to drop-off)

Administrative issues

- What can you afford?
- How long do you have?
- Do you have trained / motivated research staff?

Constructing the Survey

Five aspects of survey construction:

- Question content
- Question / response format
- Question placement
- Reducing common errors
- [Training your interviewers / enumerators]

Many aspects of survey construction are pretty commonplace -- BUT

"There is nothing so unnatural as the commonplace"

[A Case of Identity]

Constructing the Survey

● **Question content**

● Question / response format

● Question placement

● Reducing common errors

● [Training your interviewers / enumerators]

Question Content

- Too many surveys ask too many questions and don't ask them well

  Ask yourself:

- Is this question necessary? (if not you are imposing on the respondent's time)

- Is the question double-barreled? (if yes you will find it hard to interpret answer)

- Is the respondent able to answer (accurately)? (if not you will get inaccurate answers)

- Is the question loaded / leading? (if yes you will likely bias the respondent)

Think: What is your goal in formulating questions?

- To get good responses
- To get many responses
- To be sensitive to offensive or confidential issues

FOR

- Ease of analysis
- Clarity of analysis
- Relevance to intended audience
- Decency towards respondent

Constructing the Survey

● Question content

● **Question / response format**

● Question placement

● Reducing common errors

● [Training your interviewers / enumerators]

Nominal Response Format

- **Dichotomous response** format
  Do you own a mobile phone?
    □   yes   □   no

- **Nominal response** format
  Which of these best describes your religious affiliation?
    1. Catholic
    2. Jewish
    3. Muslim
    4. Secular…etc
    [numbers don't mean anything; categories should be <u>exhaustive</u>]

- Or you could have **multi-option variables**
  Please check if you have the following items at home
    1. Refrigerator
    2. TV
    3. Solar panel
    4. etc…
    [for ease of subsequent analysis you could code these as 0 if not checked and 1 if checked]

- You can use any of these formats as part of a detailed survey from which you derive correlations

Ordinal Response Format

Please rank these problems from 1 to 5 in your community (1 is most critical)

- Unemployment
- Bad water / sanitation
- Crime
- Access to important news or information
- Child health

(This style is known as a *showcard; categories need not be exhaustive; common to randomize the order of these over several respondents*)

- Ranks are meaningful but not the intervals between 2 ranks

Interval Response Format (1)

Bi-polar scale

Access to computers and the internet is one of the most important tools of poverty alleviation. Do you:

- Strongly disagree
- Disagree
- Don't know (the *neutral option*)
- Agree
- Strongly agree

- Intervals are meaningful; you structure the format so the intervals are comparable
- MUST have the neutral option

Ratio format

Count variable or ratio format:

How many children under 5 years with diarrhea did you see this month?

- None
- 1
- 2 – 5
- 6 – 10
- 11 – 20
- more than 20

- Very common "official" format e.g. DHS; LSMS; NSS…
- There is an absolute zero; ratios are meaningful
- Categories should not overlap
- Naturally lends itself to quantitative analysis

Constructing the Survey

- Question content
- Question / response format
- **Question placement**
- Reducing common errors
- [Training your interviewers / enumerators]

Question Placement (1)

How do you **order** your questions? Hard to generalize

Four main considerations:

- The answer may be influenced by prior questions (this is one reason why you either (1) don't change question order in a structured survey or (2) deliberately randomize the order for specific questions)
- If you ask a sensitive question too early the respondent may not be ready to answer (trust)
- If you ask important questions too late the respondent might get tired (or bored)
- You may have to prepare warm-up questions or relevant information on some topics

SO...

Question Placement (2)

- Explain the purpose of the survey
- Open with easy non-threatening questions
- Ask about one topic at a time (in "modules")
- Lead into different topics by explaining that you are now switching subjects
- Ask if the respondent is now willing to talk about your potentially sensitive topic
- Pre-test the sensitive questions
- Treat "pseudo-opinions" with great caution!

Most important slide of this lecture

1. TRY FOCUS GROUPS OR MINI-ETHNOGRAPHY *BEFORE* DESIGNING YOUR SURVEY.
2. PILOT YOUR SURVEY: TREAT THE FIRST FEW SURVEY ATTEMPTS AS LEARNING EXPERIENCES AS YOU FINALIZE YOUR SURVEY INSTRUMENT.
3. PILOT ONCE MORE.
4. REALLY, I MEAN IT.

Constructing the Survey

- Question content
- Question / response format
- Question placement
- **Reducing common errors**
- [Training your interviewers / enumerators]

Survey errors

- What are potential error sources? [other than poorly framed and poorly ordered questions]
- What can we do to reduce these?

  HOW to handle sources of error must balance <u>costs against harm</u> of not addressing error

Sampling error

Most common example is sampling error

(this is the major source of error that we can estimate -- the others we mostly can't)

(that's if you have a random sample or its variants)

You'll always have *random* variance wrt % of respondents from eligible sample, or in # who refuse to answer specific questions

Bias

Bias occurs when measurement is consistently above or below true population value

May be intended or unintended

Example from MICS handbook:

"Would you take your child to a clinic when he has difficulty breathing?"

(Mother will likely say "yes")

Isha Ray | MSRI-IISc Summer School 2010

Reducing bias thru' question wording

Instead ask: when your child has a cough / cold, what signs would prompt you to take him to a clinic?

(Tick off all items that the mother / respondent mentions from a list you have prepared. Don't forget to include "other" as an item)

Unit response error

If respondents are systematically different from non-respondents or drop-outs, we have unit response error

Q: How will we know?

A: We won't

So:

We might look at indirect indicators, e.g. Census data: are any groups under-represented in our sample, and are they answering differently? (Can you "augment" your sample if you see this bias?)

Reducing unit response error

There is only one way:

Maximize response rate.

-- Explain why participation is important and

-- Call back / follow up

Response rate: interviews as % of eligible sample

(this should be in final report)

Cooperation rate: (interviews as % of sample) / (completed interviews + refusals)

(this is for ongoing checks of survey quality)

Item response error

If specific questions are not answered or are answered incoherently we have item response error

If there is no pattern to the missing data items, OR if the missing data rates are low (e.g. 5%) ignore them and just move on

If lots of missing data you might need to drop some units OR some responses (and say so in the paper)

OR

You can "fill in". This is a dicey move and I usually don't like it. Especially if it is not revealed.

Non-random missing data

Most omissions are not random

Missing data usually result from difficult to answer or sensitive questions

(And don't pre-judge what is potentially sensitive)

PRE-TEST / PILOT: Your best insurance

Reducing item response error

Train yourself (or enumerator) to handle:

1. Respondent does not answer using your pre-determined categories

(What to do: Don't fill in or guess, repeat original question instead)

2. Respondent makes comment instead of answering

(What to do: Don't fill in or guess, repeat original question instead)

3. Respondent asks question instead of answering

Response is question instead of answer

Prepare ahead of time for many of these

Example: Do you think that global warming is a very serious problem, somewhat serious, not sure, not very serious, or not at all serious?

Q (from respondent): What does global warming mean?

A (from you): Whatever you understand by it

OR

A: Have short prepared answers ready, *invariant across respondents & enumerators*

(but treat potentially pseudo-opinion problem with great caution)

If respondent clearly feels she hasn't got enough info, mark "don't know"

HOW to handle sources of survey error must balance <u>costs against harm</u> of not addressing error

It's a tradeoff