

# Survey Data Collection

## Some Basics

Anthony Roman  
Center for Survey Research  
Univ. of Mass-Boston

# Goals of Presentation

- What you need to think about BEFORE you try to conduct a survey
- How to decide if this is something that may require professional assistance
- What questions to ask any vendor you may hire to conduct a survey
- How to get the “best” information to meet your needs

# What is my target population? or who do I want to learn about?

- College students – Full-time? Part-time? Grad students? Community college students? Who is in and who is out?
- People in my town – All people? Age restrictions? Voters? Taxpayers? Homeowners? Who is in and who is out?

# Sample Frame

- A list or other device used to define a population of interest.
- A sample frame must be as accurate as possible in defining your entire population of interest.

# Types of Frames

- Lists of the population – College enrollment lists, taxpayers, property owners, etc.
- Address lists within an area of interest
- All telephone numbers servicing a particular area (land lines, cell phones, portability)
- Gathering locations – homeless shelters
- Which frame best serves my purposes?

# Be Careful

- Bias – Inaccuracy created when measurements being taken do not reflect upon population of interest in some way.
- One major source of bias can be a poorly designed or inaccurate sample frame.

# Example of bias in sample frames

## Prevalence Rates

	List Strata	Remaining HHs	Overall
Chlamydia	2.30%	5.14%*	3.56%
Trichomoniasis	7.33%	9.15%	8.14%

\* difference significant at 0.05 level

# Probability Samples

- Each member of the sample frame has a known probability greater than zero of being selected into the sample. (zero implies no chance of selection)

# Non-Probability Samples

- Probability of selection of members of the sample is not known
- Sometimes called convenience samples
- Examples: Hand picking from list, TV call in samples, Stopping by a stand on a street corner

# Which do I need?

- Need probability sample if you want to make accurate estimates from your survey about your target population
- Can use non-probability sample if you want to get a idea of what may be happening in the population, but accurate estimates are not necessary

# Sampling – Types of Samples

- Simple Random Sample – everyone has equal chance of being in sample
- Complex Samples – involve stratification, clustering, unequal probabilities, etc.
- Which should I use?

# Benefits and Drawbacks

- Simple Random Samples – Best for entire population, easy to analyze and make estimates, may not be good for subgroups, may not be possible
- Complex Samples – may be necessary for subgroups, may be necessary operationally, will need specialized software for analysis

# What size sample do I need?

## Estimating a percentage of 50%

- 100 – 95% confidence interval is +/- 9.8%
- 200 -- +/- 6.9%
- 400 -- +/- 4.9%
- 500 -- +/- 4.4%
- 750 -- +/- 3.6%
- 1000 -- +/- 3.1%
- 2000 -- +/- 2.2%

# How do I collect the data?

## Interview Modes

- Face-to-Face – Interviewer administered
- Telephone – Interviewer administered
- Telephone – Computer administered (TACASI)
- Mail
- Web

# Which is best for my needs?

- Cost – Interviewer administered always more costly while web is usually the least costly
- Response rate – Interviewer administered usually gets the highest response rates while the web gets the lowest
- Sensitive questions – Computer gets more accurate responses than interviewers or mail
- Multi-mode – may be best (e.g., mail with phone follow-up)

# Interviewer vs. Computer

- Used crack in last 6 mos. 2.5% vs. 4.6%
- Paid for sex/ last 12 mos. 6.5% vs. 8.5%
- Boyfriend pushed/shoved 14.4% vs. 17.9%
- Did not have sex in 6 mos. 1.5% vs. 8.0%
- Last sex. relationship
- lasted < 6 mos. 5.8% vs. 21.3%

# Response Rates

- Percentage of eligible people who completed survey
- Should be computed using AAPOR standards, but frequently are not
- Low response rates may be a source of bias, but they also may not (may vary by question in the same survey)
- Question becomes is the variable you care about affected by bias (How do you know?)

# Creating questions

- What makes a good question or a bad question
- What characteristics of a question do you need to consider to get good information
- It is not as easy as you might think
- If possible, use questions developed for use in respected surveys (NHIS, CPS, ACS, etc.)

# Recall

## Percent Accidents Reported

Time since accident

No injury

Injury

<3 mos

94%

99%

3-6 mos

88%

90%

6-9 mos

79%

90%

9-12 mos

63%

78%

# Scale of answer categories

- How many hours per day do you usually watch television
- A.  $< \frac{1}{2}$ ;  $\frac{1}{2}$ -1; 1- $1\frac{1}{2}$  ;  $1\frac{1}{2}$  -2; 2- $2\frac{1}{2}$ ;  $2\frac{1}{2}$  +HRS
- B.  $< 2\frac{1}{2}$ ;  $2\frac{1}{2}$  -3; 3- $3\frac{1}{2}$  ;  $3\frac{1}{2}$  -4; 4- $4\frac{1}{2}$ ;  $4\frac{1}{2}$  +HRS
- % $< 2\frac{1}{2}$  HRS: A=84% B= 62.5%

# Definition of terms

- A. Do you exercise or play sports regularly?
- B. Do you do any sports or hobbies involving physical activities or any exercise, including walking, on a regular basis?

	<b>A</b>	<b>B</b>
<b>YES</b>	<b>48%</b>	<b>60%</b>
<b>NO</b>	<b>52%</b>	<b>40%</b>
	<hr/> <b>100%</b>	<hr/> <b>100%</b>

# Unfamiliar or Technical Terms

- In the past 12 months, has a doctor diagnosed you with Congestive Heart Failure?
- What was the result of your most recent A1c test?

# Imbedded Assumptions:

- In order to help you lose weight, in the next year are you planning to eat out at restaurants less often?
- Because Mitt Romney is a strong leader, do you think he should be elected President?

# Assumptions not always obvious

**ORIGINAL:** When riding in the back seat of a car, do you wear a seat belt all of the time, most of the time, some of the time, once in a while or never?

**ALTERNATIVE:**

- a) In the past year, have you ever ridden in the back seat of a car?
- b) [same as original]

	ORIGINAL	ALTERNATIVE
All of the time	30%	42%
Most of the time	17	16
Some of the time	13	8
Once in a while	8	10
Never	24	4
Don't ride in back	8	20

# Asking Multiple Questions

- When you were in the hospital, how often did your friends and family visit you?
- How often does your health care provider listen to you and spend enough time with you?

# What do you think?

Do your parents have clear rules or expectations for you about drinking alcohol?

Yes

No

# Respondent doesn't have the information

How old were you when you had your first booster shot?

How would you rate your doctor's ability to diagnose what is wrong with you?

In the past year, how many times did you get money from an ATM machine?

# Asking about other people

How much does your mother enjoy the activities in the nursing home?

How many cigarettes a day does each person in your household smoke?

Do your neighbors support the new housing bill?

# What do you think?

- Are you currently covered by any type of health insurance?

# Questions to answer before doing any survey

- Who are you interested in (specifically)?
- How can you find these people (sample frame)?
- Is this a probability sample?
- Is it simple or complex sampling? Why?
- How many people do I need to sample?
- What mode/modes will be used? Why?
- How will response rates be calculated? (get details?)
- How do I find the best way to ask questions?

[Anthony.roman@umb.edu](mailto:Anthony.roman@umb.edu)