Questionnaire Design for Social Surveys

Week 3 Asking Factual Questions

3.1 Factual Questions

Facts and Quasi Facts

Types of Questions
• Behaviors
  - use of online shopping
  - cigarette smoking
  - visits to doctors
  - exercise
  - food intake
  - sexual activities
  - voting
• Facts …

Facts and Quasi-Facts…
• Quasi Facts are Commonly seen as concrete and objective
  - with large subjective component
  - share features of factual and attitudinal questions
• Permanent
  - date of birth
  - native language
  - country of birth
• Fluctuating
  - age
  - marital status
  - income
• Quasi-Fact
  - race
  - ethnicity

Measuring Race
• What attributes define race?
• One goal: monitor and expose oppression and its results
  - are survey question the right tool?
• Variation acorns countries and time
  - U.S. Census has not used the same definition in more than two censuses
  - now collected through self/proxy identification; prior to 1960 collected based on enumerator observation
  - GB did not collect data on ethnic groups until 1991

OMB recommendations
• Federal Statistical Policy Directive #15 (1978) stipulated Federal agencies were to collect and present data on at least 4 racial groups:
  - American Indian or Alaskan Native
  - Asian or Pacific Islander
  - Black
  - White
• No reporting of multiple races
• Self-identification is preferred
• ..not tell anybody how they should classify themselves

OMG 1997
• Social definition of race recognized in this country
• Do not conform to any biological, anthropological or genetic criteria
• Include the following groups:
  - American Indian or Alaska Native
  - Asian or Pacific Islander
  - Black or African American
  - White
• Multiple race reporting is allowed

Challenges to Measuring Race and Ethnicity
• Validity of concept
• Reliability of concept: changes in self-perceptions over time
• Response selection
  - mutually exclusive categories
  - meaningful categories (for respondent)
• Comparability over time – e.g. health statistic

Questions in Context
• Context triggers cooperative conversational norms
  - Grice's 1989: participants should make their contributions useful
• Context can provide an interpretive framework (resolve ambiguities about the meaning or scope of the questions)
  - is a floor lamp counted as furniture? (schober & Conrad, 1997)
  - crime and victimization survey is just about crime
• Context can prime relevant beliefs or retrieval of relevant items
  - fear-of-crime items facilitated recall of victimizations (Cowan, Murphy & Wiener, 1978)

Why can't we just guess gender/race…? ➔ Gender Reports in CATI Surveys
• Interviewers are often asked to guess the gender of respondents in telephone surveys
• Interviewer guesses are used for a variety of purposes
  - screening to determining eligibility
  - filtering to determine survey logic
  - non-response adjustments and post stratification weighting

(Continued from above topic: Gender Reports in CATI Surveys) Research Questions & Methods
• How good are interviewers are guessing respondents gender?
  - Linguistics find pitch tones allow listener to discriminate between men and women's voices (Hess, 1983)
  - However, overlaps in pitch impede the accuracy (Ross, 1974)
  - Listeners are able to determine the sex of males easier than females (Owren 2007)
• Are there any predictors of wrong guesses?
  - language (Parris and Carey, 1996) and age (Traunmiller, 1997) of respondent
  - effects of interviewer characteristics unknown
• 28 public opinion phone surveys
  - 9/2008 – 2/2010 with total of 25,635 respondents
  - Centralized telephone facility in Poughkeepsie, NY
  - National, New York State and NYC landline samples (RDD) (cell phones excluded)
  - 475 unique interviewers complete at least one interview
  - trained interviewers are Marist College Students (ages 18-23)

Respondent Demographics (Unweighted Self-Reports)	
Age	
Under 45	26%
Over 45	74%
Race	
White	78%
African-American	10%
Other race	8%
Gender	
Male	43%
Female	56%

Interviewer Demographics	
Race	
White	77%
African-American	10%
Other race	13%
Gender	
Male	33%
Female	67%
Experience (mean number of completed surveys)	54

• How good are interviewers' gender guesses?
- Overall 8% gender guessed incorrectly
- Differential measurement error across gender groups:

Marist Poll Data	Respondent Male	Respondent Female	Total
Guessed Male	97.2%	12.6%	49.5%
Guessed Female	2.6%	86.9%	50.2%
Cannot Make a Guess	0.1%	0.5%	0.3%
Total	100%	100%	n=25,138

➔ Easier to guess male than female

• Any predictors of wrong guesses?
- multivariate model to include interviewer covariates, control for potential confounders, and to get correct standard errors
- hierarchical linear probability model included the following (clustering of interviewers taken into account):
  - dependent variables: error between guess and interviewer report: 0 = no error; 1 = error
- independent variables
  - gender of respondent
  - age of respondent: standardized continuous measure
  - race of respondent
  - gender of interviewer
  - race of interviewer
  - total number of completed surveys by interviewer (proxy for experience)
  - level of the interviewer
  - household selection
  - geographic region of respondent (tested but not included in the final model)

	Guess Error	Standard Errors
Respondent Race Black	.024***	(.005)
Hispanic	.002	(.007)
Asian	.023**	(.010)
Race Other	.011	(.009)
Respondent Standardized Age	-.006***	(.001)
Respondent Gender	.336***	(.004)
Interviewer Gender	.006	(.008)
Interviewer Race Black	.012	(.013)
Hispanic	-.002	(.018)
Asian	.025	(.036)
Other	.003	(.017)
Interviewer: Supervisor	.010	(.012)
Interviewer Experience: Level 2	.015	(.009)
Experience: Level 3	.032**	(.011)
Experience: Level 4	.044**	(.014)
HH Selection: Single person HH or youngest male	.050***	(.009)
Youngest male comes to the phone	-.343***	(.004)
Constant	-.348***	(.016)
Observations	24,188	
Sigma_U	.072	* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Sigma_E	.221	
Rho	.096	

➔ In particular, higher levels of interviewer experience were associated with higher guessing error
➔ no main effects for interviewer gender and race
➔ additional models including interaction effects showed significant effects:
- females are more likely to be miscoded by female interviewers than male interviewers
- African Americans are more likely to be miscoded than non-blacks when being interviewed by a non-black interviewer
- African American interviewer-respondent pairs have a higher probability of being miscoded than white/white pairs

Memory and Recall: Reporting Behavioral Frequencies

Memory and Recall Examples
• Specific events
- did it occur?
- when did it occur?
• Counts of activities
- How many?
- Frequency: as rate "how often during the last year did you….?"
• Since January 2013, have you looked for a job? Ask … about a specific event (comprehension) OR about past

behavior (recall) OR to aggregate across events (estimation)

Encoding Information
• Information has to be considered "salient" or "distinctive" to enter long term memory
• Salience is individually defined
  - how important is it to the respondent?
  - does it stand out from other activities?
• Deeper encoding affects accessibility
• Not encoded => not in memory at all
• Impact: underreporting
• Implications for Questionnaire Design
  - using alternative data sources instead (shot card, bills, receipts)

Reporting Immunization after visit to doctor

	Under-Report	Over-Report	Net Error
Hepititis B	51.7%	20.0%	-41.4%
DTP	41.4%	16.7%	-31.4%
MMR	33.3%	19.4%	+17.1%

Recall Methods
• Different methods of retrieval
  - free recall
  - cued recall
  - recognition
• Free recall is the most difficult (most effort)
• Recognition provides easiest, but potentially the most narrow, methodology
• Extent of cueing can have dramatic effects on the information retrieved

What affects retrieval? Forgetting
• What affects forgetting
  - the older the event, the more likely to forget
  - the less salient or the more mundane the event, the more likely to forget
  - the more repetitive/routine, the more likely to forget the individual event
• Implications for Questionnaire Design:
  - give roe time on task
  - use examples or cues
  - use shorter reference period
  - use event history calendar
    - takes advantage of structure of autobiographic structure

Time yourself
– dropped a class as an undergraduate?
– taken any antibiotics?
– not worn a seatbelt while driving?
– Have you been to the movie theater in the last month?
– How old were you when you got your first kiss?
– How many times have you seen a doctor last year?

➔ might be subject to "comprehension/interpretation" issue

Memory and Recall: Length of Reference Period

• Cannell and his colleagues did a study looking at reporting of hospital visits; record check study – had hospital records to compare to survey reports

Impact of forgetting: increases with elapsed time

Weeks between discharge and interview	**% Not Reported (n in records)**
1-10 weeks	3 (114)
11-20 weeks	6 (426)
21-30 weeks	9 (459)
31-40 weeks	11 (339)
41-50 weeks	16 (364)
51-53 weeks	42 (131)

Annual Victimization Rates by Recall Interval (per 100 persons 12+)

Type of Crime	Recall Interval 6 mos.	3 mos.
Total personal crimes	12.8	15.5
Crimes of violence	3.5	4.3
Crimes of theft	9.4	11.2
Total household crimes	23.0	26.4
Burglary	8.5	9.7
Larceny	12.7	15.1
Auto theft	1.8	2.1

Dating Events

• Date tags not stored with other information about event
• Few landmark events, for which we know dates
  - relative v.s. absolute dates
• Autobiographical sequences – memory is not a continuous linear record
• storage of date information
  - calendar or calendar-like (season)
  - socially defined period
  - elapsed time
  - Idiosyncratic reference points

When are Estimation Strategies used

• The more **frequent** the behavior, the more likely estimation will be used
• The more **regular** the behavior, the more likely estimation will be used
• The more accessible the relevant memory, the less likely estimation will be used
• Note that rate information is stored in memory for regular behavior
• Also…
  - Increase number of similar events, decrease probability of recalling any one of those events
  - reliance on generic rather than episodic memory
    - errors of omission
    - use of schema for reconstructing

Memory and Recall: Comprehension and Retrieval

How are objects defined?

• Retrieval from memory is dependent on what R understands what is to be searched
• How does object line up with memory traces?
  - "meta-memory" – feeling of knowing
  - what are relevant memory clusters?

Context of Retrieval

• What does questionnaire define as the object?

- introduction to items
- surrounding items
- wording of question itself

• use introductory statements to facilitate accurate recall

Setting the context for victimization surveys

• Introduction to National Crime Survey:
- "Please tell me about any crime that may have occurred in the last 6 months"

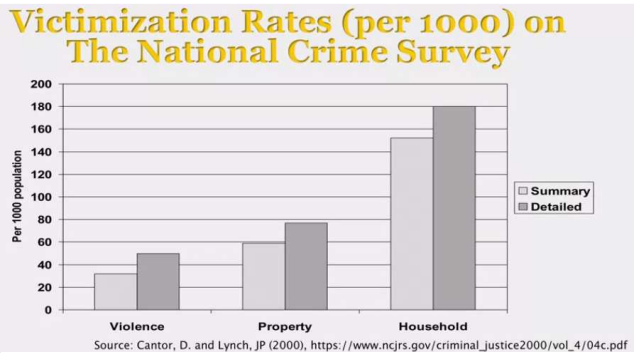
• Introduction to National Violence Against Women Survey:
- "We are particularly interested in learning about violence women experience, either by strangers, friends, relatives or even husbands and partners"

Detailed Cueing defines objects

• Use of cues focuses retrieval on just those objects
- cues might interfere with retrieving other information (blocking)
- cues implicitly exclude objects that are not mentioned

Summary v.s. Detailed Cueing

Summary	Detailed Cueing
Did you have your pocket picked?	Was something belonging to you stolen, such as –
Did anyone steal things that belonged to you from any car or truck, such as packages or clothing?	Things that you carry, like luggage, a wallet, purse, briefcase, book Clothing, jewelry or calculator Bicycle or sports equipment
Was anything stolen from you while you were away from home, for instance at work, in a theater or restaurant, or while traveling?	Things in your home – like a TV, stereo, or tools
Was anything at all stolen from you during the last 6 months?	Things outside your home such as a garden hose or lawn furniture Things…



3.2 Sensitive Questions

Asking Sensitive Questions

Learning Outcomes

• What makes a question sensitive?
• What are the consequences of sensitivity?
• What can we do to minimize the consequences?
• What conclusions can we draw from practice?

Are these questions sensitive?

Do you….

• Almost always wear a seat belt when you drive or ride in a car?
• avoid certain foods because they are high in fat or cholesterol?
• smoke at least one cigarette a day?
• drink three or more caffeinated beverages a day?
• spend 1/2 hour or more in moderate or strenuous physical activity at least 3 times a week?
• have a library card?

Types of Sensitive Questions

• Private information
- income
- identifier (SSN, name, address, telephone #)

• Illegal behaviors
- drunk driving
- drug use

• socially stigmatizing behavior/opinion
  - abortion
  - # of sexual partners
  - discrimination

Mixture of Issues
• Can the information be used in a legal way?
• Is the information personal?
• Does the information relate to the image of the respondent?

Different Level of Sensitivity
• Question is threatening **to ask**, regardless of respondent's answer
  - e.g. income
  - respondents will refuse to answer
• Question is threatening **to answer**
  - e.g. drug use
  - respondent may give the wrong answer
  - depends on what the actual behavior is

Consequences
• High unit nonresponse rates
  - people are less likely to participate in surveys with sensitive questions
    - people with undesirable behavior / attitudes less likely to participate
• High item nonresponse rates
  - people are more likely to not provide a response when asked sensitive questions even after they agreed to participate in the surveys
    - ~25% of item nonresponse rates to income questions
• Misreporting

Why do people edit?
• "Norm" in the 'uncertainty space' of sensitive questions
  - good citizens vote!
  - people should not use drugs!
  - it is considered irresponsible to drive while drunk
• People choose to give the 'norm' answer
  - to avoid embarrassment
  - out of privacy/confidentiality concerns
• People are more likely to edit and misreport
  - if they have something to hide
  - if topics are getting more sensitive
  - if (certain) bystanders are around

Factors
• Mode of data collection
  - self-administered modes elicit more socially undesirable and less socially undesirable behaviors/attitudes than interviewer-administered modes
    - presence of interviewer
• Interview setting
  - bystander presence (parents, sibling, spouse, other)
    - if bystanders do not know the answer or R afraid of letting bystanders know, R more likely to edit and misreport
• Wordings of the questions

Asking Sensitive Questions: Techniques

Context
• Response format is important

• Set the context of the question
  - embed question within related set of questions
  - best to embed question to communicate behavior is acceptable
• Setting context with prior questions
  - provide a context that implicitly suggests the behavior in desired way
    - permissive context for undesirable behavior
    - restrictive context for desirable behavior
• example from Tourangeau and Smith on Abortion Questions
  - Q1) In general, would you say that you drink more than your friends less than your friends, or about the same amount as your friends?
  - Q2) Think the friend you know who drinks the most. About how many drinks would you say that person usually has?
  - Q3) And how about you? On days when you have any alcoholic beverages, about how many drinks do you have?

Ease the Task
• Familiar words
  - critics suggest that standardized wording makes interview artificial
  - emulate normal conversation
  - variable wording??? Impact on measurement, Interviewer variance
• level of detail requested (e.g. income categories; response scale)

Proxy Reports
• Data from informants
  - tradeoff between loss of quality v.s. social desirability issues
  - proxy knowledge
  - dependent upon relationship between reporter and individual for whom proxy information is obtained

Deliberate Loading
• Deliberate loading of question
  - can be used to both reduce over reporting of desirable behavior (voting) and underreporting of undesirable behavior
• embed threatening topic into less threatening topics
  - reduce perceived importance
  - threat is determined, in part, by context
• Deliberate Loading for Undesirable Behavior
  - everybody does it
    - even the calmest parents get angry at their children some of the time. Did your children do anything in the past seven days to make you angry?
  - assume behavior and ask about frequencies (presupposition)
    - how many cigarettes do you smoke a day?
    - danger of insulting those who do not engage in behavior
  - use of authority
    - wine has recently been shown to reduce cholesterol levels and improve digestion…. Followed by questions of interest
• Deliberate Loading for Desirable Behavior
  - casual approach: "did you happen to"
  - reasons why not: voting, seat belt use

"have you ever" Questions v.s. Current Questions
• Undesirable behavior
  - recent behavior is most threatening
  - ask "have you ever" Question before current behavior Question
• Desirable behavior
  - more threatening to admit never doing something than to admit within a short time frame or more recently not doing something
  - tradeoff between social desirability and sufficient observations for rare events

Asking Sensitive Questions: Indirect Techniques

RRT – Example
• Tails – Question A
• Heads – Questions B
A: Did you get a parking ticket last month?
B: Is your mother's b day in June?

	R to all	Est. R. unrelated Question	Inferred R. to target Q	R to t-Q as pop. %
Yes	20%	4%	16%	32%
No	80%	46%	34%	68%
	100%	50%	50%	100%

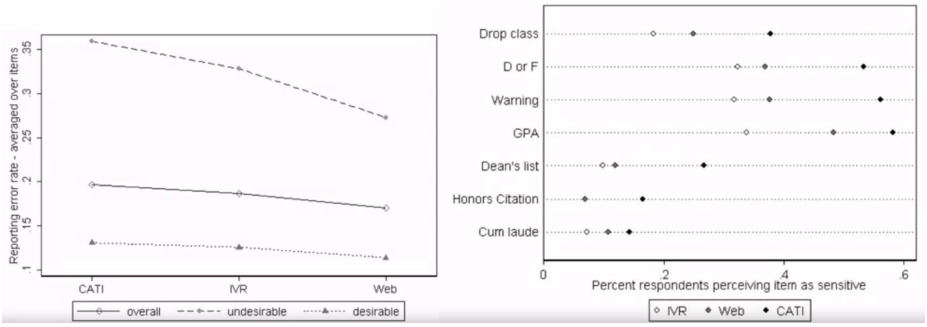
Mode, Privacy, and Confidentiality

Mode
• Mode effects on social desirability
  - SD bias stronger in interviewer administered surveys
  - assessment usually without record data
  - little knowledge about relative effects between self-administered modes
• Effects on desirable and undesirable items
  - underreporting for undesirable behavior, over-reporting for desirable
  - studies mostly done for undesirable behavior
  - little knowledge for relative effects of desirable and undesirable behavior
• Socially desirable items and sensitivity
  - social desirability (sensitivity) of items is not a fixed parameter
  - SD could depend on 'true' score
  - SD could depend on perceived sensitivity

Study Design
• JPSM 2005 practicum
• sample of alumni
  - graduating classes 1989 – 2002
  - sample drawn from administrative records
  - academic record data available
• socially undesirable and desirable behavior
  - failing grades, academic probation, dropping a class
  - honorable mention, GPA
  - donations/donation '04; amount donated '04
  - alumni association
• mode of administration
  - initial contact by phone with random assignment to
    - web
    - CATI
    - IVT

	CATI	WEB	IVR
Completes	320	363	320
% complete	31.9	36.2	31.9
RR1	18.8	11.3	12.1

➔ <u>expect least accurate responses to sensitive questions in CATI survey</u>



➔ (left) error difference between CATI and Web largest for "undesirable" questions
➔ (right) items considered most sensitive in CATI

Set up the Interview Properly
• Assure the respondent that information will be kept private
  - consider tradeoffs between anonymous and confidential
• Assure the data will be protected so identities of respondent will be protected

Confidential v.s. Anonymous
• Confidential – the identity of respondent is known (or knowable), but information will be restricted to authorized project members
• Anonymous – no one knows identity of respondent. Cannot connect identity of respondent to survey responses
• Is this anonymous?
  - list sample of respondents is drawn from school roster. An ID number is given to each student. The respondent fills out a survey with no name on it. The ID is printed on the corner of the survey

Human Subjects Consideration: Legal Obligations
• The extent that information can be protected depends on legal standing of survey
• Data are subject to legal requests (e.g. subpoena) unless there is special protection
  - legislative protection
  - certificate of privacy
• A data security plan should cover:
  - who will have access to the data
  - how data will be secured
    - electronic
    - paper
  - how data will be stored and/or destroyed after study is completed
• Plans for public released of information