Introduction to Survey Methodology

Professor Ron Fricker
Naval Postgraduate School
Monterey, California

Reading:
Dillman et al. chapters 1 & 2
Contact Information

• Professor Ron Fricker
  – Phone
    • Comm: 831-656-3048
    • DSN: 756-3048
    • Fax: 831-656-2595
  – E-mail: rdfricker@nps.edu

Come by or call anytime!

3/26/13
Professional Experience

• Academic credentials
  – Ph.D. and M.A. in Statistics, Yale University
  – M.S. in Ops Research, The George Washington University
  – B.S. in Mathematics from the United States Naval Academy

• Teaching credentials
  – Started teaching post-graduate courses in mid-80s
  – Have taught at NPS, RAND Graduate School, and USC

• “Real world” credentials
  – Former active duty naval officer
  – Senior Statistician specializing in survey and military research at the RAND Corporation

• Can find out more at http://faculty.nps.edu/rdfricke/
Course Schedule

Wednesday, May 15
0800-0930: Introduction to Survey Methodology – Dillman et al. chapters 1 & 2
1000-1130: Crafting Good Survey Questions – Dillman et al. chapters 4 & 5
1230-1400: From Questions to Questionnaire – Dillman et al. chapter 6
1430-1600: In-class Exercise: Designing a Short Survey

Thursday, May 16
0800-0930: Summary of Instrument and Question Design Guidelines
1000-1130: Evaluating Survey Questions (and Instruments) – Fowler chapter 5
1230-1400: Implementation (Fielding) Strategies – Dillman et al. chapter 7
1430-1600: In-class Exercise: Expert Review & Cognitive Interviewing

Friday, May 17
0800-0930: Elements of the Sampling Problem – Schaeffer et al. chapter 2
1000-1130: Simple Random Sampling – Schaeffer et al. chapters 3 & 4
1230-1400: Stratified and Cluster Sampling – Schaeffer et al. chapters 5 & 8
1430-1600: In-class Exercise: Introduction to R

Tuesday, May 21
0800-0930: Postcollection Processing of Survey Data: Coding & Weighting – Groves et al. chapter 10
1000-1130: Postcollection Processing of Survey Data: Imputation – Groves et al. chapter 10
1230-1400: Introduction to Survey Analysis
1430-1600: In-class Exercise: Postcollection Processing of Survey Data

Wednesday, May 22
0800-0930: Categorical Data Analysis for Survey Data – Lohr chapter 10
1000-1130: Factor Analysis – Fricker, Kulzy & Appleget Phalanx article
1230-1400: Regression for Survey Data – Lohr chapter 11
1430-1600: In-class Exercise: Illustrative Survey Analyses

3/26/13
Course Materials

• Textbook: *Internet, Mail, and Mixed Mode Surveys* by Dillman et al.
  – Also excerpts from:
    • *Survey Methodology* by Groves, et al.
    • *Improving Survey Questions: Design and Evaluation* by Fowler
    • *Elementary Survey Sampling* by Schaeffer, Mendenhall, Ott, and Gerow
    • *Sampling: Design and Analysis* by Lohr

• Lecture notes with all slides & reading excerpts
• Software: R

➢ All notes on-line at
  http://faculty.nps.edu/rdfricke/Survey_Short_Course.htm
Course is All About Collecting and Analyzing Survey Data

• In this short course you’ll learn:
  – How to design good survey questions and craft an effective survey instrument
  – About the trade-offs between various survey modes
  – The rudiments of sampling for surveys
  – Some statistical methods useful for analyzing survey data
Good Surveys Require Skills in Design, Sampling & Analysis
Also Requires Designing the Survey from Multiple Perspectives

Respondent’s Perspective
• Why should I take this survey?
  o Why is it worth my time?
  o Are the surveyors competent?
• Can I (easily) take the survey?
  o Are the questions clear?
  o Are there reasonable responses to choose from?

Sponsor’s Perspective
• Will the survey answer my research objectives?
  o Are the necessary topics covered in the survey?
  o Are the right people being surveyed?

Analyst’s Perspective
• Given the questions, their responses, and the sample, are the sponsor’s questions answerable?
  o Are the questions (analytically) clear and do they map to the objectives?
  o Are the response sets clear and the categories mutually exclusive?
  o Is the difference between a neutral, no opinion, and don’t know response clear?

Good Surveys
Goals for this Lecture

• Define what we mean by the term “survey”
  – Characteristics of the types of surveys this class will focus on
  – Distinction between “polls” and “surveys”

• Learn some basic survey terminology

• Discuss the basic steps in conducting a rigorous research survey

• Discuss the various survey modes
  – Pros and cons
  – Trade-offs
What is a Survey?

• A survey is a:
  – systematic method for gathering information
  – from (a sample of) entities
  – for the purposes of constructing quantitative descriptors
  – of the attributes of the larger population of which the entities are members
What’s the Difference Between a Survey and Simple Data Collection?

• Surveys are just one form of data collection – but they are unique in that:
  – Data gathered via a process of question asking and answering
  – Data is often “generated” at the time the question is asked

• Hence, data accuracy very much depends on the survey design and the survey process
  – For example:
    • Does the survey design minimize human error?
    • Does the survey process engender the good will and motivate those taking the survey?
Focus of Course (and Text)

• There are lots of types of surveys
• We will focus on surveys where
  – Information is primarily gathered by asking people questions
  – Information is collected by either:
    • Interviewers asking questions and recording responses (i.e., “interviewer assisted” surveys)
    • Respondents reading and recording their own answers (i.e., “self-administered” surveys)
  – Information is only collected from a subset of the population – a sample – rather than from all members
The Goal of Surveying

• Collect information about a population
• Such information should be
  – Accurate and
  – Generalizable

as well as
  – Relevant and useful (to answer some question)
  – Unavailable by other means
For the Results to be “Accurate” and “Generalizable”...

2. The survey respondents must have characteristics similar to those of the larger population.

1. Answers people give must accurately describe their characteristics

From Questions to Characteristics

Survey Questions

27. Please tell me whether you would be willing to accept someone from a different **ethnic group** as…
   a. A coworker  
   b. A neighbor  
   c. A close friend  
   d. A relative by marriage

29. Please tell me whether you would be willing to accept someone from a different **religious group** as…
   a. A coworker  
   b. A neighbor  
   c. A close friend  
   d. A relative by marriage

Classify an individual as tolerant or not
From Sample to Population

Unobserved population statistic → Sample statistic

Inference
Polls vs. Surveys

• There is no clear distinction between the two terms
  – “Poll” most often used for private sector opinion studies
    • Use many of the same design features as studies that would be called surveys
  – “Poll” rarely used to describe government or scientific surveys

• To me, the term poll implies either
  – a commercial or less-scientific study, or
  – a quick turn-around survey whose results may be of short-term interest
Survey Terminology

- **(Survey) Instrument.** The entire survey questionnaire, including instructions, questions and response scales

- **Item.** A question and its associated response scale in an instrument
  - In practice, the terms “item” and “question” often used interchangeably

- **Mode.** How survey participants are contacted, questions are administered, and responses are recorded
  - “Contact mode,” “response mode,” “follow-up mode”
Survey Terminology (continued)

• **Closed-ended question.** A survey question in which the response scale is a fixed set of pre-defined responses
  – **Likert scale.** Closed-ended response scale that allows the respondent to state how much they agree or disagree with a statement

• **Open-ended question.** A survey question designed to allow respondents to respond to the question in their own words
Survey Terminology (continued)

• **Validity.** The extent to which a survey question accurately measures the property it is supposed to measure

• **Bias.** A systematic error in the data
  – **Non-response bias:** units that do not answer your questions look different than those that do
  – **Selection bias:** units with a particular trait and/or strong opinions are favored
    • Strong opinion
    • Access to survey mode (telephone, internet, etc.)
  – **Sensitivity bias:** answers to questions of salary/sex/other social taboos might not be truthful
Survey Terminology (continued)

- **Population.** The group of people (or organizations) whose activities, beliefs or attitudes are being studied.

- **Sampling Frame.** The list of people (or organizations) drawn from which the sample will be drawn.
  - Explicit sampling frame: An actual list
  - Implicit sampling frame: A method for enumerating the population in order to select a sample.
Survey Terminology (continued)

- **Sample.** The set of people drawn from the sampling frame
  - **Census.** A survey in which every member of a population is invited to respond
- **Respondent.** An individual providing answers to survey questions
- **Response rate.** The number of completed surveys divided by the number of eligible potential respondents contacted
Traditional Survey Modes

• Mail
  – Paper questionnaire sent to respondents
  – Self-administered and mailed back

• Telephone
  – Interviewers call respondents on telephone
  – Interviewer-assisted

• In-person
  – Interviewers go to respondent’s home or office
  – Interviewer-assisted
Options Have Multiplied in the Age of Computers and Telecom

Figure 5.2 The evolution of survey technology.

Mode Considerations

• Modes differ by:
  – Degree of interviewer involvement
  – Degree of interaction with respondent
  – Degree of respondent privacy
  – Channels of communication
  – Technology usage

• Considerations:
  – Cost
  – Timeliness
  – Accuracy
Broadly speaking…

- Interviewer-assisted surveys more costly than self-administered
  - Face-to-face most expensive, then telephone, then mail, then web
- Interviewer-assisted surveys less prone to various errors and item nonresponse
  - Particularly compared to self-administered paper survey
  - More or less compared to self-administered computer-assisted surveys depending on sophistication of program
- Self-administered less subject to sensitivity bias
  - Can combine some self-administered questions into a predominantly interviewer-assisted mode
However, timeliness is not just about…

- …the length of time a survey is in the field
  - Really, it’s from instrument design through analysis
    - Much of that is not affected by survey mode
  - Compared to a mail survey, web-based surveys save on two main things:
    - Mailing time to send out and return surveys
    - Time required to do data entry, coding, and cleaning for paper surveys
    - Yet, it can also take a lot of time to code, test, and de-bug very complicated web (more generally, computer-assisted) survey programs
And, cost is not just about…

• ...manhours
  – On mail surveys, printing & postage costs can be significant
    • 1,500 surveys x 4 mailings at $1.50 each = $9,000!
    • 1,500 $1 incentives = $1,500
  – On telephone surveys must consider fixed costs of phone
    and CATI equipment + variable costs
  – On face-to-face interviews, travel costs can be significant
    (not to mention interviewer time)

• And in terms of manhours, don’t forget:
  – instrument design time;
  – in non-electronic modes, data entry and coding time;
  – non-response follow-up efforts; and
  – analysis time
And accuracy is not just about…

- … response rates and missing items
  - Remember total survey error – what mode or modes will decrease best all types of error
- Lots of considerations
  - Frame coverage
  - Nonresponse rates
  - Sensitivity bias, etc…

Steps in Conducting a Survey – Part I

1. Clearly define research objective(s)
   – Write it down and make sure all parties agree!

2. Write an analysis plan

3. Decide on survey mode(s)
   – How will you contact potential respondents?
   – In what media will the survey be given?
   – How will you follow up with non-respondents?

4. Determine fielding strategy
   – What methods will you use to maximize response rates?

5. Design the survey questions and the instrument
   – Make sure the questions support the objective(s) and the analysis plan
     • Otherwise, don’t ask it!

3/26/13
Steps in Conducting a Survey– Part II

6. Determine sampling strategy and sample size
7. Obtain Institutional Review Board (IRB) and/or other approval as necessary
   – Are respondents promised anonymity? Confidentiality?
   – What is the impact if their survey responses become known?
8. Pre-test, pre-test, pre-test
   – Give the survey to some test subjects and get their feedback
     • What works and what doesn’t?
     • Are you getting correct data/information?
   – Revise and re-pre-test as necessary

9. Iterate steps 1-8 until all bugs and issues are wrung out and worked through
Steps in Conducting a Survey—Part III

10. Draw sample and field the survey
   - Follow up with non-respondents

11. Assemble the survey data
   - Clean data as necessary

12. Analyze the data and report results
   - Summarize the data
     - Weight as necessary and appropriate to infer back to population
     - Calculate and report margin of error
   - Evidence of bias?
     - Unit and item non-response

13. Communicate survey results to respondents (particularly if promised)
One View of the Steps

Figure 2.4 A survey from a process perspective.
Another View of the Steps

Stage 1: Planning and Development of Survey
- Collect background data for planning survey design
- Prepare questionnaire outline
- Plan preliminary operations
- Develop preliminary analysis plan and report outline

Stage 2: Pretest
- Prepare sampling frame
- Select pretest sample
- Draft questionnaire
- Prepare preliminary codes
- Hire subcontractor
- Pretests (2)

Stage 3: Final Survey Design and Planning
- Develop sampling plan
- Revise questionnaire
- Prepare final questionnaire
- Revise survey and design operations plan
- Revise analysis plan; Draft final report outline

Stage 4: Implementation of Survey and Data Collection
- Select sample
- Establish sample control
- Prepare final questionnaire
- Coordinate and manage project with subcontractor
- Collect data
- Reduce data: Editing, Coding, Data entry, Cleaning
- Check data quality: Verification, Validation
- Prepare raw data file

Stage 5: Data Coding and Data-File Construction
- Collect data
- Reduce data: Editing, Coding, Data entry, Cleaning
- Check data quality: Verification, Validation
- Prepare raw data file

Stage 6: Research and Analysis of Data
- Analyze data
- Draft report
- Prepare final report

Adapted from Czaja and Blair (1996)

Excerpted from What is a Survey? by Fritz Scheuren
Other Data Collection Methods

• Surveys are not the only (nor necessarily the best) way to collect data

• Other methods include
  – Administrative records
  – Focus groups and qualitative investigations
  – Randomized experiments

• Which is “best” depends on the research question(s) and/or the purpose for which the data will be used
Why Good Survey Methods Are Important
“On another note, I probably should have taken the opportunity to take your Survey class back at NPS. …One thing I would tell any Marines and Army that may be out there and due to graduate soon, surveys are quickly becoming THE data source out here. As we pull out of different areas, quality reporting continually decreases and soon survey data (good or bad) will become the main gauge we have on what is going on in the country. SigActs will still happen of course but we just may not know about nearly as many of them.”

Capt Joe Maddux
What We Have Covered

• Defined the term “survey,” including the
  – Characteristics of the types of surveys this class will focus on
  – Distinction between “polls” and “surveys”
• Defined additional basic survey terminology
• Discussed the basic steps in conducting a rigorous research survey
• Discussed the various survey modes
  – Pros and cons
  – Trade-offs