

36-303: Sampling, Surveys and Society

Components of a Survey
Brian Junker
132E Baker Hall
brian@stat.cmu.edu

Handouts

- Graded Quizzes (Your Score is on p. 2)
- Today's Lecture Notes
- HW01

Outline

- Quiz Results
- Wrap up previous lecture
- Process of Conducting a Survey
 - Defining Research Objectives
 - Mode of Data Collection; Target Population; Frame
 - Measurement; Errors of Observation
 - Sample; Errors of Non-Observation
 - Coding, Editing and Post-Survey Processing
 - Analyzing the Data, Writing the Report

Quiz Results

■ Scores:

6 | 5
7 | 04
7 | 5
8 | 02344444
8 | 6677799
9 | 00000000000333333444
9 | 777
10 | 00

median

■ It was an easy quiz

- 80 or above
 - Generally feel pretty good
 - Errors were sloppy or minor
- Below 80 – a significant chunk is missing
 - Median/Outliers
 - Histogram/Boxplot
 - Confidence Interval
 - Scatterplot
 - Summation Notation
 - Expected Value
 - Binomial Distribution

Quiz Results (Cont'd)

- Most answers pretty obvious – ask your friends or check with us
- CI for Mean Test Performance...

N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
93	82.61	1.06	10.21	58.00	77.00	84.00	91.00	99.00

$$\text{StDev} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 10.21$$

$$\text{SE Mean} = \text{StDev} / \sqrt{n} = 1.06$$

$$\begin{aligned} 95\% \text{ CI} &= \text{Mean} \pm 1.96 \times (\text{SE Mean}) \\ &\approx (82.61 - 2 \times 1.06, 82.61 + 2 \times 1.06) \end{aligned}$$

Quiz Results (Cont'd)

$$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- What is role of i ?
- What is role of x_i ?
- How do we calculate it?
- What is it?

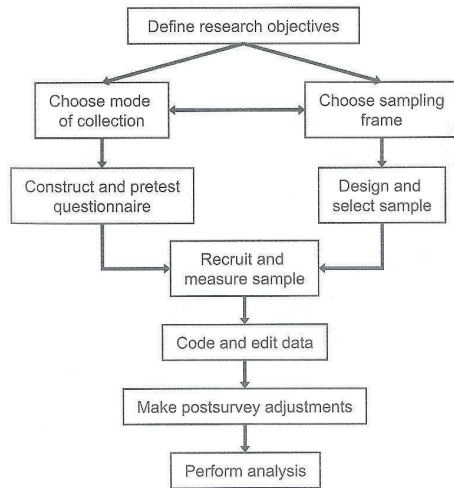
Wrapping up Previous Lecture

- Elements of a Sample
- Does Sample Represent Population?
- Non-sampling errors and Sampling Errors
- What can we say about
 - Population of Interest
 - frame/list
 - sampling technique
 - sample size
 - response rate
 - mode of interview
 - possible sources of selection bias and inaccuracy
 - other details of methodology relevant to our inferences

Team Assignments; Project Outline

- Team assignments were emailed out Monday afternoon
 - As the projects get underway there may be some small adjustments in some teams
- Next deadline: Tue Jan 25: ***Propose two topics!***
 - (Indicate interest/non-interest in Jewish Chronicle project as well)
- (HW01 is also due Tues Jan 25).

Process of Conducting a Sample Survey



18 January 2011

9

Defining Research Objectives

■ Research Question(s)

- Is it of interest? (*Who Cares??*)
- Can it be answered with available methods?
- Can a survey on it be conducted and analyzed within budget (\$\$, time, effort, irritation, ...)?
- Surveys are not well-suited to cause-effect questions (Why not? Think about 36-309...)

■ Target Population (*This is harder than it sounds!*)

- What population is relevant to the question?
- What population can you construct a good sampling frame for?

■ Construct (*What information do you seek?*)

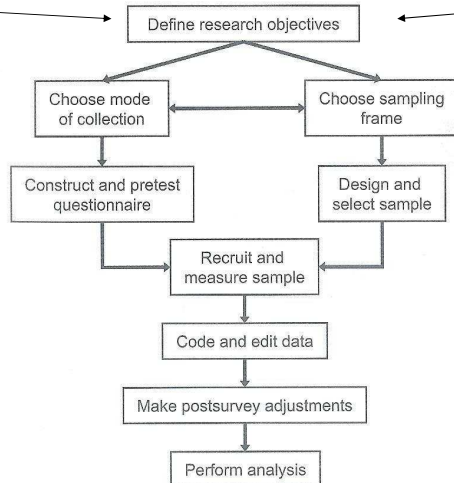
- "Number of jobs created in last month"
- "Consumption of beer in the last month"
- "Knowledge in mathematics of eighth grade school children"
- "Optimism about one's financial status"

18 January 2011

10

Process of Conducting a Sample Survey

- Question
- Construct



18 January 2011

11

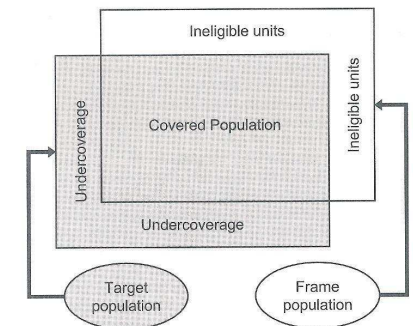
Mode of Data Collection and Sampling Frame

■ Why Sampling Frame ≠ Target Population?

- Population may not have a natural frame
- Mode of data collection may restrict frame

■ Mode of Data Collection

- Interview
 - Face to face?
 - Telephone?
- Self-report
 - Face to face?
 - Internet?
- Direct
 - Administrative records?
 - Observe prices, soil samples, type of nbhd, etc.

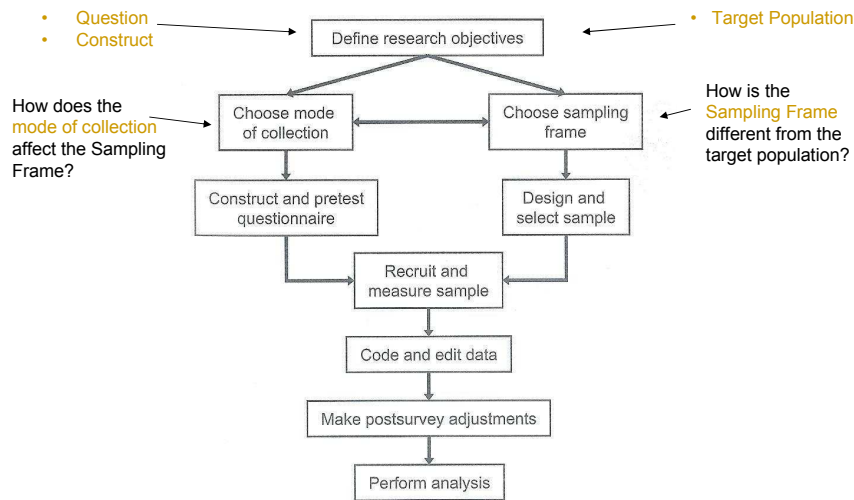


Coverage Error – the extent to which the **Sampling Frame** does not cover the **Target Population**

18 January 2011

12

Process of Conducting a Sample Survey



18 January 2011

13

Measurement; Response; Errors of Observation

- **Measurement:** How we gather information for constructs
 - ❑ Chemical analyses of soil samples
 - ❑ Electronic measures of traffic flow
 - ❑ Observations of classroom teaching
- **Questions** posed to respondent are common
 - ❑ Oral (face-to-face interview)
 - ❑ Visual (self-report or computer-assisted interview)
 - ❑ Based on some stimulus (reaction to watching a video, listening to music, reading a story)

18 January 2011

14

Measurement; Response; Errors of Observation

- **Responses** depend on the form of the question
 - ❑ Multiple choice
 - ❑ Fill in the blank
 - ❑ Longer user-generated response
- **Nonresponse**
 - ❑ Didn't understand, didn't see, or refused question (**item nonresponse**)
 - ❑ Not home, not approached by interviewer, refused phone call, etc. (**unit nonresponse**)

18 January 2011

15

Measurement; Response; Errors of Observation

- **Errors of Observation (Measurement Error)**
 - ❑ Deviations of measurement from underlying construct
 - ❑ Inaccurate measurements
 - Inaccurate administrative records
 - Poor chemical analysis of soil
 - Untrained interviewers/observers
 - Memory/attention/understanding/truthfulness of respondents
 - ❑ Item Nonresponse

18 January 2011

16

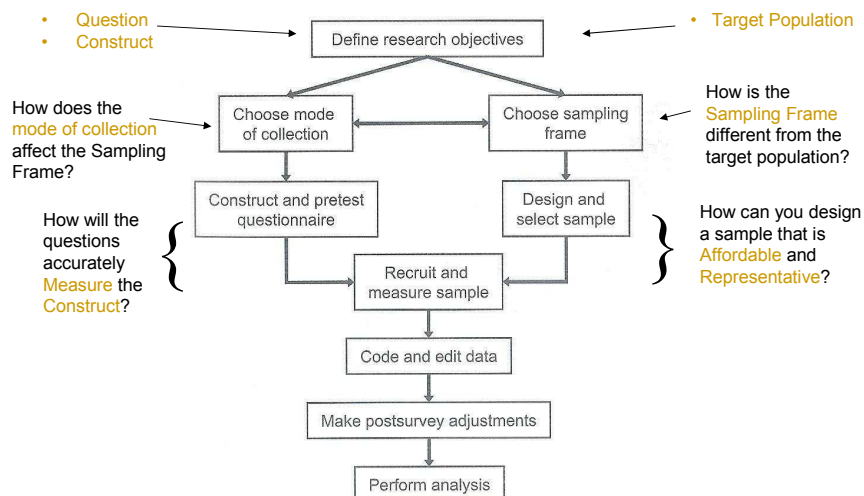
Designing a Sample; Errors of Non-Observation

- We want to design a sample that is
 - **Affordable** (time, money, effort, accessibility...)
 - **Representative** (of the frame? Of the target population?)
- Simple populations with good frames
 - Simple sample designs and analyses suffice
- Complex populations or poor frames
 - **Stratified sampling** and **Clustered sampling** common
 - More complex designs require more complex analyses
- **Followup** for Unit Nonresponse?

Designing a Sample; Errors of Non-Observation

- **Errors of Non-Observation**
 - Deviations between the sample and the target population.
 - How representative of the Sampling Frame is the Sample?
 - How representative of the Target Population is the Sampling Frame (Coverage Error...)
 - How do we followup unit nonresponders?
 - Sample more units to replace them?
 - Keep after them until they respond?

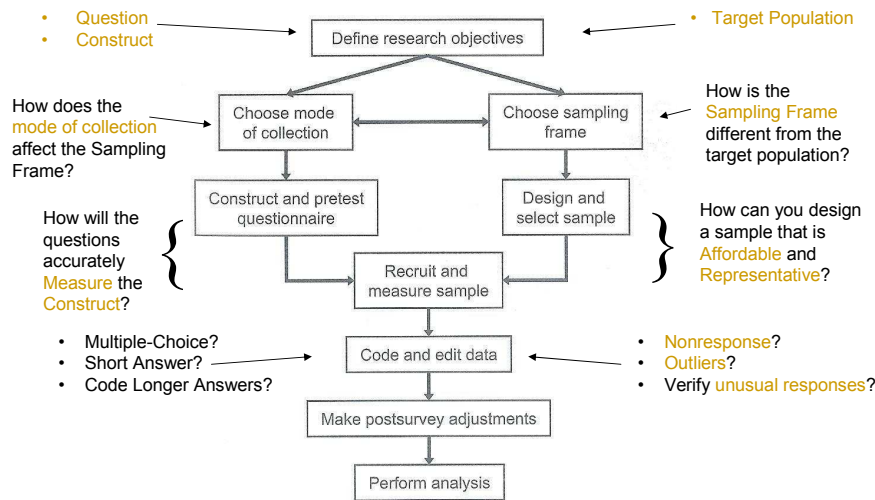
Process of Conducting a Sample Survey



Coding and Editing Data

- **Coding depends on measurements**
 - Multiple choice?
 - Fill in the blank, long-answer, taped conversation?
 - Accuracy of chemical analysis?
- **Nonresponse**
 - Unit nonresponse? Successful Followups?
 - Item nonresponse? Refused? Not asked? Not reached? Not understood?
- **Outliers**
 - What is an outlier?
 - Include anyway? Drop?
 - Followup to verify value?
- **Inaccurate Data**
 - Detection? Followup? Correct value? Drop case?

Process of Conducting a Sample Survey



18 January 2011

21

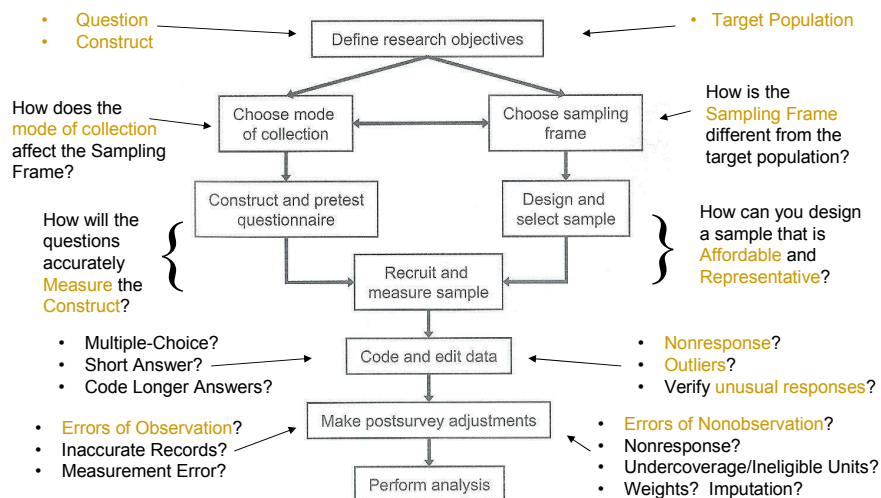
Post-Survey Adjustments

- Adjustments for
 - Patterns of unit nonresponse (did women respond less than men?)
 - Under- or over-coverage of the sampling frame (no phone numbers for homeless men?)
 - Inaccurate or outlying data, ...
- **Weights** (only 20% of sample was women but 50% of population are women, so “weight up” women by 5/2)
- **Impute** missing values (unit nonresponse and item nonresponse)

18 January 2011

22

Process of Conducting a Sample Survey



18 January 2011

23

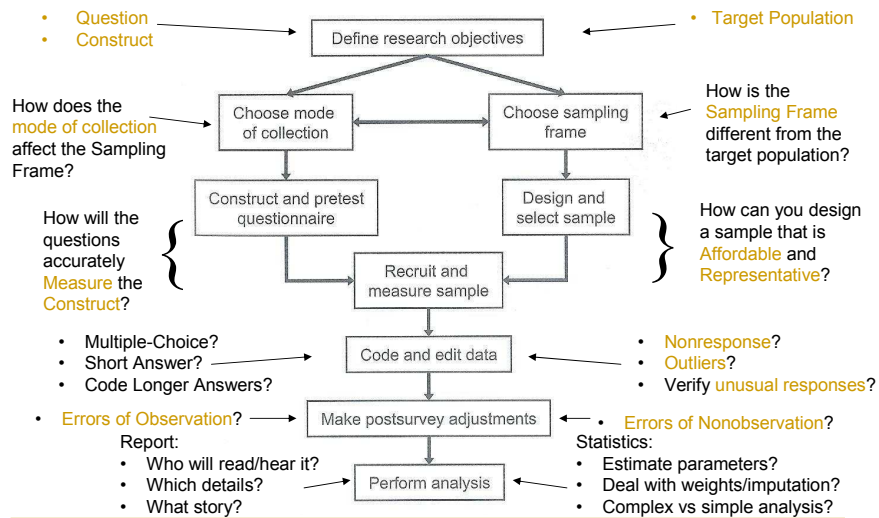
Performing the Analysis

- Statistical analysis
 - What quantities should be estimated? How?
 - Do we have to deal with weights? Imputation?
 - Simple designs can use simple statistics; complex designs require complex statistics
 - Statistics cannot fix (or even quantify!) all errors
- Report writing
 - Who will read the report? **How** will they read it?
 - How much detail is needed? Where should it go?
 - What is the interesting story you are trying to tell?
 - Research objectives: Who Cares???

18 January 2011

24

Process of Conducting a Sample Survey



Review

■ Quiz Results

■ Team Assignments

■ Process of Conducting a Survey

- What are the various components of a survey?