
36-303: Sampling, Surveys and Society

What is Sampling?
Brian Junker
132E Baker Hall
brian@stat.cmu.edu

19 January 2012

1

Review Quiz

- Fill in your name.
- Answer questions on the handout.
- You have 30 minutes.

19 January 2012

2

Handouts

- Quiz
- Project Ideas
- Lecture Notes
 - Quiz
 - Forming Project Groups
 - Project Ideas and Class Schedule
 - What is Sampling?

19 January 2012

3

Forming Project Groups

- Find people you can work with
- Use the discussion board Blackboard to find a group to join or find a person to add to your group
- Groups should be 4-5 students each
- Email brian@stat.cmu.edu with your proposed group members, by Monday at 5:00pm. ONE EMAIL PER GROUP.
- I will assign you to a group if you do not choose.

19 January 2012

4

Project Ideas, Class Schedule

- Project Ideas
- Handouts next week will give details on
 - Schedule of topics in lectures for the course
 - Schedule of deadlines/landmarks for your projects.

What is Sampling?

- Sampling is a statistical process of “purposefully” selecting a subset of units from a population in order to make inferences about the entire population.
- Every sampling method is designed to make errors (because not everyone should be in the sample!)
- A census is designed to make no errors (because everyone should be in the census!)

A Sample and a Census are not the same thing!

- In a sample there are two sources of error
 - The data set is not the whole population
 - Errors in the design or implementation of data collection
- In a census there is only one source of error
 - The data set is (supposed to be) the whole population
 - Errors in the design or implementation of data collection

Sampling in the US Census

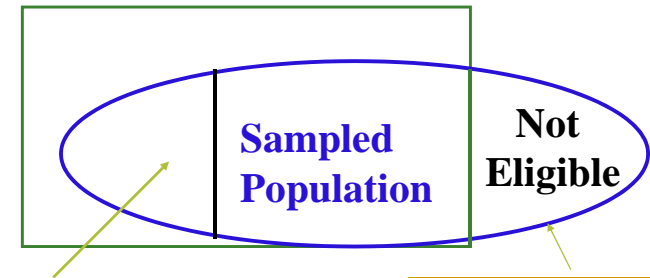
- Constitution requires "actual enumeration" (Census) for the purpose of apportioning US Representatives.
- US Dept of Commerce v US House of Representatives (1999) : Cannot use additional sampling to correct census errors.
- Utah v Evans (2003): “Hot deck” imputation (a method for correcting data collection errors in samples) can be used to correct census errors.

Elements of a Sample

- Key elements to understanding properties of sample include:
 - *Target Population* - collection of observations we want to study (e.g. possible voters in NH).
 - *Sampled Population* – all possible observation units that might have been sampled.
 - *Sampling Frame* – list of all sampling units (e.g. list of telephone numbers).
 - *Sample* - subset of population.
 - *Sampling Unit* - unit we actually sample (e.g. household).
 - *Observational Unit* - element to be measured (e.g. individual).

Schema

Target Population



**Refusals, Unreachable,
Incapable, etc.**

**Sampling Frame
Population**

Does Sample Represent Population?

- “Representativeness” comes from
 - (a) match between target population and sampled population.
 - (b) method for drawing sample.
- Two kinds of errors:
 - Non-sampling – *can be reduced by careful design of the survey*
 - Sampling – *can be quantified by statistics, reduced by increasing sample size*

Two Kinds of Errors

- Non-sampling errors:
 - **Selection bias** - part of target population is not in sampled population.
 - **Measurement bias** - measuring instrument has tendency to differ from true value in one direction.
- Sampling error - results from taking a sample instead of whole population.
 - The method of sampling determines whether & how statistics can be useful in quantifying (and reducing) sampling error!
 - The “best” way to select a sample by using probability methods, because this gives us a valid statistical basis for inference.

Methodological Features of Examples

- 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 on methodology relevant to inferences of interest

Summary of Today's Lecture

- Review Quiz.
- Project groups
- Project proposals
- Key elements of sampling
- What makes a sample representative?