

Sampling

Feb 13/06

Outline

1. Introduction
2. Bias
3. Sampling Error (Non-Sampling Error)
4. Types of Sampling
 - a. Simple Random
 - b. Systematic
 - c. Stratified Random
 - d. Non Random (convenience/snowball)
5. Response Rates

1. Introduction

Your experience...

Were you part of a sample?

THINK BREAK:

Have you ever sampled anything?

Jam, fudge, free massage?

How do you know it represents the whole?

2. Bias

In order to generalize results from a sample it has to be... representative..

A non-representative sample is biased.

Three types of sample bias:

- Not random
- Sampling frame is inadequate
- Non-response

3. Sampling Error

So you remove all bias from your sample, do you remove error?

Sampling error =

“Differences in data between a random sample and the population from which it is selected.”

Get out your money!

Terms Review:

Population = all the units from which the sample is to be collected.

Census = Data collected from all units in a population

Sample = segment of population selected for analysis; sub-set of population

Sample frame = the listing of all units in the population from which the sample will be selected.

Non-Sampling Error

= differences between a sample and the population that arise either from deficiencies in the sampling approach or from other validity problems (see below)

-Happens through:

Sample bias (Replication)

- not random
- poor sample frame
- non-response

Validity problems

- poor questions
- poor interviews
- flawed data processing

THINK BREAK

What have I learned?

What's a bit muddy?

What would I ask?

Can I think of other examples of populations, censuses and samples?



4. Types of Samples

- a. Simple Random
- b. Systematic
- c. Stratified Random
- d. Multi-stage Cluster Sampling

- e. Non-probability Samples
(convenience/snowball/quota)

a. Simple Random Sample

- Most basic form of probability sample
- Each unit of the population has an equal probability of inclusion

E.g. Hamilton voting intentions

b. Systematic Sample

Version of Simple Random

Selecting units directly from the sample frame
without using random numbers

E.g. Hamilton Voting Intentions

- $100,000 / 500 = 2000$
- Choose a random number between 1-2000
 - E.g. # 344 becomes first case
 - Second case is $344+500 = 844$
 - Then 1344, 1944, 2444, and so on...
- Assumes no order in the list

c. Stratified Random Sampling

Stratifying the population by criteria and then doing random (or systematic) sampling from within each of the individual strata or groups

E.g. Hamilton voting intentions

2-3 different ridings within the Greater Hamilton Area

d. Multi-stage Cluster Sampling

To avoid costs one creates clusters from which one will choose randomly those to be interviewed

E.g. Canadian Voting Intentions (from Text)

Non-Probability Samples

Any form a choosing sample participants which do not include random selection allowing one to generalize to a population

Convenience sampling=choosing cases that are accessible to the researcher

Snowball sampling=multi-stage approach of choosing accessible interviewees and then using their contacts to gain access to others

Quota sampling=choosing those cases accessible but in relative proportion to certain population categories

5. Response Rates

Response rate = The percentage of people from the originally-generated sample who complete the survey.

Different types of non-response

- Not available or able
- Refusals
- Break-offs or Incompletes

Debate over the denominator

Effects of Non-response – non-sampling error that can effect the generalizability

p.133 in text

Reducing non-response:

- Call-backs to when people are available

- Confidence/Self-assurance

- Clarify not SUGGING

- Dress for success

Large number of break-offs or incompletes:

- Poor questionnaire

- Poor interviewers

Lowest response rates for online surveys!

- No longer novel

- Hate unsolicited e-mails (or block them)

THINK BREAK

What have I learned?

What's a bit muddy?

What would I ask?



Summary

1. Introduction
2. Bias
3. Sampling Error (Non-Sampling Error)
4. Types of Sampling
 - a. Simple Random
 - b. Systematic
 - c. Stratified Random
 - d. Non Random (convenience/snowball)
5. Response Rates

Tutorial

Why are response rates so low these days? Does this undermine all social science and market research?

How do I get my head around the fact that a 1000 person sample of Americans is just as robust as a 1000 person sample of Canadian, even though the population of the USA is ten times larger?

How do media reports of social and market research tend to talk about the sample? Is it significant? What do they pick up as being important? What important things do they miss?