

CHAPTER 12 NOTES

SAMPLE SURVEYS

In 1976, feminist Shere Hite published *The Hite Report*, in which she suggested that a great many women are unsatisfied with their loves lives. Hite Gathered the opinions expressed in her books through two polls. For the first she mailed out approximately 100,000 questionnaires. She got back 3,019 questionnaires (3%) and felt justified generalizing on that basis. Hite believed that such a large response justified her conclusions.

Dissatisfied with Marriage	98%
Had extra-marital affair	75%

What are a few potential problems with this report?

IDEA 1: TAKING A SAMPLE

We wish to know about an entire **population** of individuals, however, examining all of them is typically impractical, if not impossible. Therefore, we will examine a **sample** of individuals selected from the population.

IDEA II: RANDOMLY SELECT THE SAMPLE

Randomizing makes sure that, on average, the sample looks like the rest of the population.

Parameters and Statistics

Definition: In statistics, a **parameter** is a numeric quantity, usually unknown and unknowable, that describes a certain population characteristic.

Definition: A **statistic** is a quantity, calculated from a sample of data, used to estimate a parameter.

IDEA III: SAMPLE SIZE MATTERS

The fraction of the population you sample is unimportant, it is the **sample size** itself that determines the precision of the estimate

Identify the **population, parameter, sample, and statistic.**

The Gallup Poll interviewed 1007 randomly selected U.S. adults aged 18+, March 23-25, 2007. Gallup reports that when asked when (if ever) the effects of global warming will begin to happen, 60% of respondents said the effects had already begun.

NOTATION

Below are examples of the notation for parameters and statistics.

	Parameter	Statistic
Mean		
Standard Deviation		
Proportion		
Correlation		

Statistics are usually represented by Latin letters with other symbols (e.g. \bar{x} , s_x , p , r , etc.)

Parameters are normally represented by Greek letters (e.g. μ , σ , π , ρ , etc.)

SIMPLE RANDOM SAMPLE (SRS)

In a simple random sample **each combination** of individuals has an equal chance of being selected.

How to obtain a SRS:

- Define the **sampling frame**
- Assign each individual in the sampling frame an identification number, and obtain random digits to determine what individuals are selected.

USING THE RANDOM NUMBER TABLE

- Step 1: Figure out how many digits you need
- Step 2: Decide what direction you are going to travel
- Step 3: Find a starting place on the table
- Step 4: Travel in the direction you’ve already chosen, writing down the digits you need, ignoring any spaces in the table

A P P E N D I X E Tables

Row	TABLE OF RANDOM DIGITS					
1	96299	07196	98642	20639	23185	56282
2	71622	35940	81807	59225	18192	08710
3	03272	41230	81739	74797	70406	18564
4	46376	58596	14365	63685	56555	42974
5	47352	42853	42903	97504	56655	70355
6	20064	04266	74017	79319	70170	96572
7	73184	95907	05179	51002	83374	52297
8	72753	36216	07230	35793	71907	65571
9	03939	30763	06138	80062	02537	23561

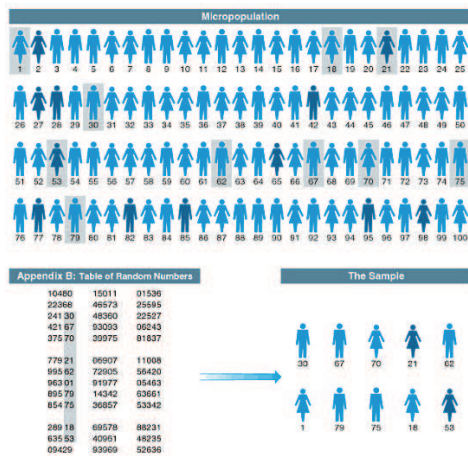


Image from *The basics of social research* by Earl R. Babbie

EXAMPLE OF SRS:

What Proportion of Students are in Class today?

Row	TABLE OF RANDOM DIGITS					
1	96299	07196	98642	20639	23185	56282
2	71622	35940	81807	59225	18192	08710
3	03272	41230	81739	74797	70406	18564
4	46376	58596	14365	63685	56555	42974
5	47352	42853	42903	97504	56655	70355
6	20064	04266	74017	79319	70170	96572
7	73184	95907	05179	51002	83374	52297
8	72753	36216	07230	35793	71907	65571
9	03939	30763	06138	80062	02537	23561

STRATIFIED RANDOM SAMPLING:

- Method

- Benefits

Example

120 **men** and 80 **women** are in company and we want their opinions on policy of arrival of children. How should we sample 20 people (10% of population).

CLUSTER SAMPLING

- Method

– Sometimes we cannot get a sampling frame for the units we are interested in studying.

Example: Tree Killing Bark beetles have decimated approximately 4 million acres of forest in the northwest. How could we perform a survey of the trees to estimate what proportion are infected by the beetle?

SYSTEMATIC SAMPLING:

- Assumptions

Frame can be **ordered**

- Method

Select every k^{th} individual to be in the sample, starting at a randomly selected point in the ordered sampling frame

$k = 1$	$k = 2$	$k = 3$	$k = 4$
Amy	Joe	Laura	Corey
Mary	Anne	Steven	Shirley
Tanner	Wiley	Liam	Logan

MULTI-STAGE SAMPLES:

13

VOLUNTARY RESPONSE BIAS

Situation

A large group of individuals is invited to take a survey, and all who respond are counted.

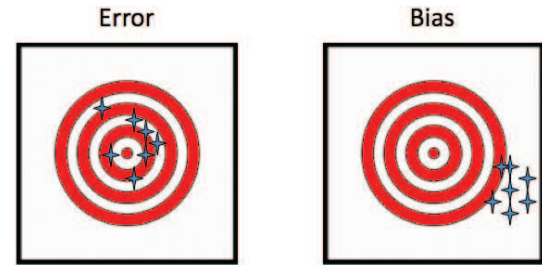
Problem

Example: The Hite Report where Hite sent out 100,000 questionnaires.

Category	Hite	ABC News
Dissatisfied with Marriage	98%	7%
Had extra-marital affair	75%	7%
Total Questionnaires	100,000	1,875
Responses	3,019	1,505
Response Rate	3%	80%

15

What can go Wrong?



- Sampling methods that tend to over- or underemphasize some characteristics of the population are said to be **biased**.
- Error can be explained and quantified

14

CONVENIENCE SAMPLE BIAS

To obtain a **convenience sample** we simply include a group of individuals that is easy to sample.

Problem

Example

We wish to determine what ISU students think about the department of residence. To do this, we go door to door in a dormitory, and ask questions to determine their opinion.

Example

A Company at the mall asks shoppers who walk by to fill out a survey rating their store.

16

UNDERCOVERAGE BIAS

Problem

Some portion of the population is not sampled or has a smaller representation in the sample than it has in the population.

Example

1. Literary Digest poll in 1936 predicted the republican candidate would win big over FDR in the Presidential election. Literary Digest had over 2 million respondents, however, the frame was a list of people with a telephone or car.
2. If your sampling frame was the list of all people who voted in the midterm elections for a poll for the presidential election.

17

NONRESPONSE BIAS

Problem

Those individuals who do not respond to a survey are somehow different from the respondents.

Example

1. A congressman sends out mailings to constituents asking their opinions of the economic recovery effort.
2. Telephone surveys conducted between 9 am and 5 pm

18

RESPONSE BIAS

Response bias refers to any problem in the survey design that influences the responses.

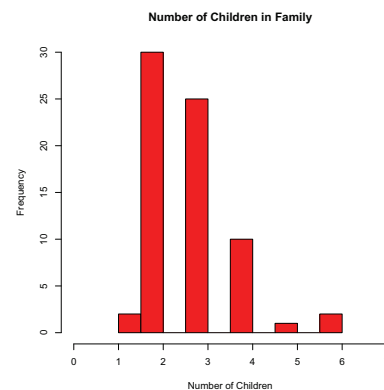
Example

- Wording of a question
 - *Do you favor or oppose gun control?*
 - *Would you favor or oppose a law that would take away your constitutional right to bear arms?*
- Asking about illegal activity

19

Wacky Samples cont.

1. We are interested in determining the average number of children per family.
2. I asked the quesiton, “How many brothers and sisters are in your family, including yourself?”



20

Wacky Samples cont.

Los Angeles Times, April 20, 1993: “Poll finds 1 in 3 Americans open to doubt there was a Holocaust.”

Question: “Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?”

Not Possible didn’t happen	66%
Possible didn’t happen	22%
Don’t know	12%

Survey repeated a year later with “Does it seem possible to you that the Nazi extermination of the Jews Never happened, or do you feel certain that it happened?”

Certain it happened	92%
Possible never happened	1%
Don’t know	8%

Example:

Identify: Population, Sample, Statistic, and Parameter.

An airline surveys 25 passengers on a trans-atlantic flight to see what proportion will want coffee when they wake up. There are 250 passengers on the flight.

Example: Name each sampling method described below.

An airline wants to survey a sample of the 250 passengers on a trans-atlantic flight.

- Pick every 10th passenger from a list of ticket numbers starting at a random ticket number.
- Sample the first 25 people to board the flight.
- From the Boarding list, randomly choose 25 seat numbers and survey the passengers sitting there.
- Randomly choose 5 people from First Class and 20 people from Lower Class.