

How Do We Obtain Data?

Sampling Techniques

J. Garvin



Slide 1/13

Sampling Techniques

Data is often collected via surveys or questionnaires, distributed to many people.

It is seldom feasible to survey *every* individual about a particular issue for several reasons:

- Cost (very expensive)
- Time (need a lot)
- Complexity (lots of data)

Smaller groups of individuals may be surveyed, and their responses used as *predictors* for everyone.

J. Garvin — How Do We Obtain Data?
Slide 2/13

Samples and Populations

A *population* is the set of all individuals that belong to a group being studied.

For example, a researcher is performing a study involving migration habits of polar bears in Canada. The population is the set of all polar bears in Canada.

A *sample* is a subset of a population.

In the example above, the researcher might tag 100 polar bears to make up a sample.

J. Garvin — How Do We Obtain Data?
Slide 3/13

Samples and Populations

When choosing a sample, it is important to select individuals from appropriate groups.

If inappropriate choices are made, then the sample may not be representative of the population. This is a form of *bias*.

When conducting surveys or questionnaires, be sure to identify the appropriate population.

Example

What is the population being targeted by the question “Who is your favourite teacher at Cawthra Park?”
The population is all Cawthra Park students, as they would have experience with the teachers here.

J. Garvin — How Do We Obtain Data?
Slide 4/13

Samples and Populations

Your Turn

Identify the population being targeted in each survey question.

- “What brand of snowboard performs best?” All people that snowboard.
- “Which political leader won the debate?” All people that watched the debate.
- “Which course did you prefer: history or geography?” All people who took both history and geography.

J. Garvin — How Do We Obtain Data?
Slide 5/13

Common Sampling Techniques

Simple random sampling is probably the most basic form of random sampling.

Simple Random Sampling

Every member of the population has an equal probability of being part of the sample.

The inclusion, or exclusion, of any individual does not affect the probability of any other individual being selected.

For instance, a Data Management class of 30 students makes a population. A simple random sample of five individuals may be created by randomly selecting five names from the attendance sheet.

J. Garvin — How Do We Obtain Data?
Slide 6/13

Common Sampling Techniques

A more structured approach is used in *systematic sampling*.

Systematic Sampling

Selecting individuals at regular intervals, determined by both the population size and the desired sample size.

$$\text{interval size} = \frac{\text{population size}}{\text{sample size}}$$

Using the Data Management class of 30, a systematic sample of five individuals may be created by randomly selecting one name from the first six students on the attendance list, then counting every sixth person thereafter.

Common Sampling Techniques

Sometimes, it is important to ensure that data is collected proportionately from several groups.

Stratified Sampling

The same *proportion* of individuals is selected from two or more groups that share common characteristics.

Common characteristics may include:

- Age
- Gender
- Income
- Health

Common Sampling Techniques

Example

SAC is planning a school-wide dance, and wants input from all grades. They plan to give surveys to 20% of all 1500 Cawthra Park students.

Since each group is to be proportionally represented, the table below shows the number of students surveyed in each grade.

Grade	Students Enrolled	Students Surveyed
9	500	$500 \times 0.2 = 100$
10	400	$400 \times 0.2 = 80$
11	350	$350 \times 0.2 = 70$
12	250	$250 \times 0.2 = 50$

Other Sampling Techniques

Cluster Sampling

Randomly selecting groups that are thought to be representative of the population.

For example, a car manufacturer might survey all of the dealers at randomly-selected dealerships to obtain information regarding car sales.

Multi-Stage Sampling

Uses two or more levels of random sampling.

For example, to obtain information regarding television viewing habits, a corporation might randomly select 100 cities across Canada, then select 1000 households from each of these 100 cities to give questionnaires.

Other Sampling Techniques

Voluntary-Response Sampling

Allows any member of a population to participate.

Responses from voluntary-response samples are often skewed, because those who choose to participate are often not representative of the population.

Convenience Sampling

Samples members of a population that are easily accessible.

This is similar to cluster sampling, without the random selection process.

Sampling Techniques

Your Turn

Identify the sampling method in each scenario.

- A politician's office calls every 100th number in a telephone directory to ask voters their opinions on an issue. Systematic sampling.
- A store's receipt includes a link to a web-based survey soliciting customer opinions. Voluntary-response sampling.
- You ask the three students nearest to you about the solution to a problem. Convenience sampling.