

Sampling Methods

J. Garvin



Slide 1/14

Hypotheses

Much of the time, researchers spend their time trying to find relationships between things.

For instance, a researcher might want to determine if there is a relationship between the amount of protein consumed in an individual's diet and their likelihood of developing heart disease.

When designing an experiment or an investigation, a researcher will probably formulate a *hypothesis*.

A hypothesis is a statement that can be tested as to whether it is true or false.

A good hypothesis should be based on evidence – an “educated opinion”, rather than a blind guess.

J. Garvin — Sampling Methods
Slide 2/14

Hypotheses

Example

Write a hypothesis about a relationship between a student's grade and the number of hours he/she spends completing homework.

A hypothesis might be “if a student completes more homework, his/her grade will increase.”

This may *not* be the case, of course. It is entirely possible that a student's grade is affected more by other variables instead.

In this case, the hypothesis would be false and the *opposite* of the hypothesis would be true.

The opposite of the hypothesis is “if a student completes more homework, his/her grade will not increase.”

J. Garvin — Sampling Methods
Slide 3/14

Sources of Data

When collecting data, researchers can draw upon different sources.

Primary sources are those that provide original data, such as experiments or surveys conducted by the researchers themselves.

Secondary sources are those that provided data gathered from others, such as journal articles, newspaper reports, or surveys conducted by others.

When conducting a study, it is important to reference all data sources using an appropriate format, especially when they are secondary sources.

J. Garvin — Sampling Methods
Slide 4/14

Sources of Data

Example

Identify each source of data as primary or secondary.

- Your conversations with local WWII Veterans.
- Data about smoking and lung cancer found on the Statistics Canada website.
- Checking the price of a new computer via the websites of a dozen electronics stores.

The conversations are primary sources, since you conducted them.

The smoking data comes from a secondary source, Statistics Canada. Others performed the research.

Since you are collecting data directly from the stores' websites, price-checking is using primary sources.

J. Garvin — Sampling Methods
Slide 5/14

Sampling Methods

Canada has a population of approximately 35 000 000 people.

Surveying all of these individuals would be difficult to organize, costly and time-consuming.

If a researcher wanted to obtain data about a particular topic, he/she might survey a *sample* instead – a smaller group of people, taken from a population.

The data would only be valid, however, if those in the sample truly represent the population.

For instance, if a sample of Canadians only included middle-aged men, then it would not accurately reflect Canada's population.

In this case, we would say that the sample is *biased*.

J. Garvin — Sampling Methods
Slide 6/14

Sampling Methods

The simplest form of sampling is *simple random sampling*.

Using this method, any individual from the population is eligible to be in the sample. Each individual is “randomly” selected.

For example, a researcher might have a list of 1000 names. He/she might use a random number generator to select 100 values between 1 and 1000.

The researcher can then select those 100 positions in the list to obtain the individuals for the sample.

Sampling Methods

Systematic random sampling is similar to simple random sampling, but is slightly more organized.

Using the previous example, a researcher might have a list of 1000 names and want to sample 100 individuals.

This means that the researcher wants $\frac{100}{1000} = \frac{1}{10}$ of the individuals on the list to be in the sample.

He/she might use a random number generator to select a values between 1 and 10. Then, beginning at that position on the list, the researcher select every 10th individual to create the sample.

Sampling Methods

When multiple groups need to be represented *proportionally*, researchers may use a *stratified* sample.

For instance, a simple random sample of the students in a high school might elicit more responses from one grade than another. In this case, the sample might not be representative of the population.

A stratified sample, however, would ensure that each grade level is represented as a percentage of the total population.

This means that a group that makes up a larger percentage of the population receives more representation than one that makes up a smaller percentage.

Sampling Methods

Example

In a high school, there are 500 grade 9 students, 300 grade 10s, 250 grade 11s and 150 grade 12s. How many students in each grade should be sampled if a researcher wishes to sample of approximately of 150 students?

There is a total of $500 + 300 + 250 + 150 = 1200$ students in the school. 150 students represents $\frac{150}{1200} = 12.5\%$ of the school's population.

Therefore, 12.5% from each grade should be randomly sampled in order to represent the population proportionally.

Sampling Methods

The totals from each grade level are below.

| Grade | Number of Students Sampled |
|-------|----------------------------|
| 9 | $500 \times 0.125 = 62.5$ |
| 10 | $300 \times 0.125 = 37.5$ |
| 11 | $250 \times 0.125 = 31.25$ |
| 12 | $150 \times 0.125 = 18.75$ |

The researcher should sample roughly 62 grade 9s, 38 grade 10s, 31 grade 11s and 19 grade 12s, for a total of 150 students.

Sampling Methods

If a sample is not representative of the population, then it is *biased*.

Bias can be accidental (e.g. randomly selecting a majority of individuals from a small subgroup) or intentional – that is, bias is introduced to skew opinions or promote a particular viewpoint.

Non-random sampling can often be convenient (e.g. sampling the first 5 tables served at a restaurant about their experience) or cheaper (e.g. surveying employees at one branch of a bank rather than across all branches).

Sampling Methods

Example

Identify the type of sampling used in each scenario.

- Surveying every 20th visitor to an online store.
- Interviewing 25% of all workers in each position at a manufacturing facility.
- Asking 5 people nearest to you on a bus for their opinion about minimum wage laws.

Since the visitors are sampled using a fixed interval, this is systematic random sampling.

The manufacturers are being represented proportionally, so this is stratified sampling.

The bus passengers have probably been selected for convenience, so this is non-random sampling.

Questions?

