

Lecture 1

Introduction to Sampling

Sampling is a process that

Two important sampling questions:

1. What

2. How

Question (1) is the _____ and (2) is the _____

The design problem involves questions about the _____

Example: Suppose the goal is to find the average age of students enrolled at MSU.

- Can we use the collected ages from our course survey, why or why not?
- Now devise a strategy to collect this information. Consider how students will be selected, how many students to question, and how to ask students their age.

By properly addressing the design problem, the analysis problem will often not be difficult. In other words, given the design, the researcher should know the format of the analysis prior to data collection.

Populations and Sampling Units

A **survey** is

A **census** is

An **observation unit** is

The **target population** is

The **sampled population** or study population

Now back to the age of MSU students. What is the target population? Discuss a situation where the target population does not match the sampled population.

It is certainly desirable for the target and study populations to match. However, this is often not the case. When they do not match, it may not be possible to make statements about the target population from data collected on the study population.

Similarly the *scope of inference* is restricted to the study population in our hypothetical data collection studies.

The potential members or units of a sample

A **sampling frame** is a

Consider three sampling schemes: one that calls dorm rooms and asks the person answering the phone their age, one administered in classes, and one that interviews students based on a master list from the registrar. What are the sampling units and sampling frame in each situation?

Note: At this point we clarify the distinction between *individuals in a population* and *sampling units in the sampling frame*.

If the sampling frame consists of individuals in the study population then the sampling units and the individuals in the study population are the same.

However, a sampling plan could consist of sampling subgroups of individuals. In this case, the sampling units are subgroups of individuals.

For many populations, the sampling unit is obvious.

For other populations, the sampling unit may not be obvious. For example, when surveying a geographical region, you may have to use a map to identify what is the basic sampling unit. Consider an experiment designed to assess the effect of pine beetle on Montana forests, how might you conduct this study?

This type of sampling introduces a number of problems, namely:

1. The
2. A
3. A

Estimates vs. Estimators

The goal of sampling is to make conclusions about some characteristics of interest for one or more populations of interest based on the data collected.

Using our class data on the number of hours spent at Bridger Bowl last winter answer the following questions:

1. What is the sampled population in this case?
2. What is the sampling frame for this problem?
3. How would you estimate the average number of hours for the class?

A **parameter** is a

A **statistic**

In general, the value of the population parameter is unknown. Statistics computed from data can provide information about the unknown parameter.

The process of estimating a population parameter by a statistic

Prior to data collection, a sample statistic is a random variable and is called a

After collecting a sample and conditional on the data, a sample statistic

Using data from our class survey, what is the point estimator and point estimate for the number of hours spent at Bridger Bowl this winter?

Later in the course we will discuss interval estimation in the form of confidence intervals to express uncertainty in point estimates.

The researcher's sampling goal is to collect a sample that is

Accuracy is related to

Sampling and Estimation Concepts

In one of the most common sampling situations, we assume the population consists of a finite number N of sampling units.

Associated with each of the N units is a

Each y -value is considered a fixed quantity representing that unit. In other words, we assume the sequence of population y -values (y_1, y_2, \dots, y_N) is fixed.

For each sampled unit there will be

A **sampling design** is

The classical sampling designs

Sampling designs that are based on planned randomness are called **probability samples**.

When taking a simple random sample (SRS) of size n , the possible samples consist of n distinct units selected from the population of N units, and $P(\mathcal{S})$ is the same for every possible sample \mathcal{S} . Thus $P(\mathcal{S}) = 1/(\text{the total number of unique samples of size } n)$

Sketch out pseudocode to sample 5 members of the class to give their course presentations on Friday.

The typical inference problems in sampling are

1. the

2. an

This variability assessment is often an

Ideally, we would like a sampling strategy which will yield samples that produce estimates with small variability that are centered around the true value.

Thus, by choosing an appropriate sampling design and estimation method, the researcher can often obtain unbiased estimates without making additional assumptions about the population.

Selection by use of probability samples removes intentional or unintentional human sources of bias (such as the tendency to select units with larger or smaller than average values). Use of probability samples to generate a representative sample is especially desirable when there are parties with conflicting interests (e.g., a fish study that will be used by fishery management, commercial users, and environmental groups).