

## Statistics

Goal: use data to draw insights about population.

Population  $\rightsquigarrow$  well defined collection of objects interested in studying.

### Example

- $\rightsquigarrow$  ① people in the USA
- ② Cars in Maryland
- ③ Televisions manufactured in a factory

Variable  $\rightsquigarrow$  characteristic of an object in the population

Quantitative



measurements, counts,

numeric values



have units

Qualitative/Categorical



Qualitative properties like color, sex, race, nationality, etc



can put into categories.

### Goal Restated:

Draw insights/inferences of properties of variables at the level of the population



### Goal Restated:

Draw insights/inferences of properties of variables  
at the level of the population



- Problems
- ① Population might be very large / intractable.
  - ② Too expensive to access all objects in population
  - ③ All objects in the population might not be accessible.



Ex: studying certain fungi in the Amazon rainforest.

### Solution:

Work with a subset of the population



sample

Generally:

size of sample << size of population

Advantages → have access to all objects in the

sample

→ relatively small set of objects

→ ease of calculations

Work with a subset of the population  
↓  
sample

Generally:

size of sample << size of population

- Advantages
- have access to all objects in the sample
  - relatively small set of objects
  - ease of calculations

General Goal



Note: There is no free lunch!

The price of working with a subset of the population

↓  
additional care in choosing the sample.