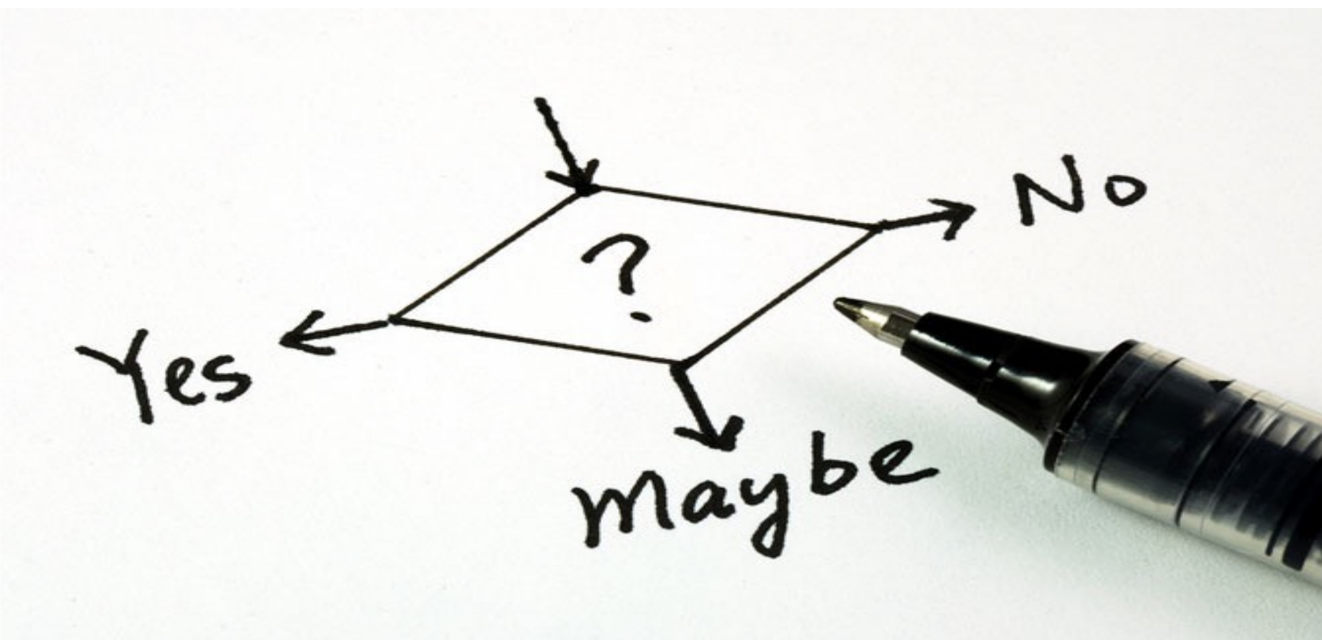
Don’t hesitate to raise you hand during lectures if you have any questions.

I’m also used to “read rooms” to see when students have questions.

Statistics is the science of assisting in decision making with incomplete knowledge



We can't measure everything: statistics are based on samples!

Biologists are relatively small: we collect smaller number of things to generalize to all things!

- From Chris Lortie

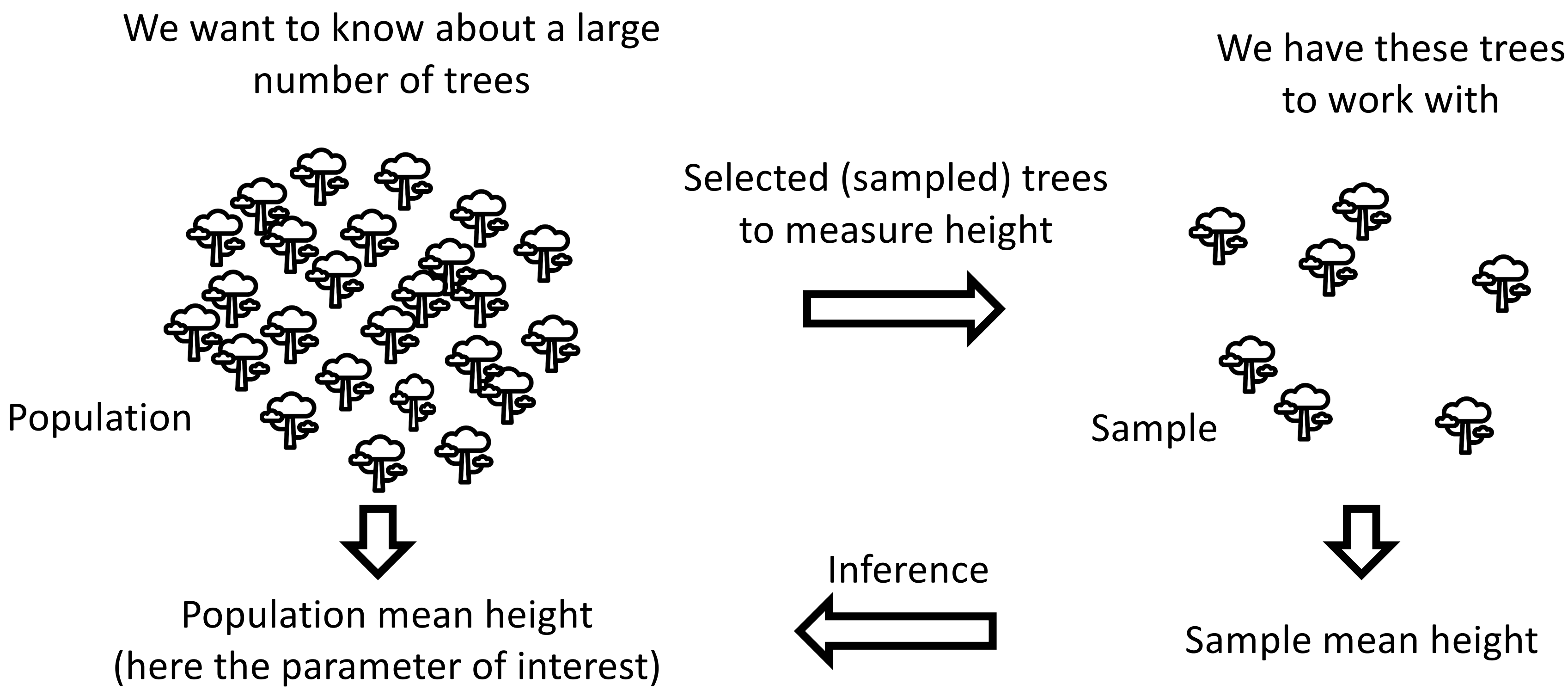
Statistics are based on samples!

The most important goal of statistics is to infer an unknown quantity (e.g., average height of a species of plant) for an entire population of plants based on sample data (a subset of observations from the population)!

Biologists are relatively small

e.g., = exemplī grātiā (for the sake of example!)

The most important goal of statistics is to **infer an unknown quantity** (e.g., height) of a population (species of tree in a forest) based on sample data!



Inspired by https://www.cliffsnotes.com/study-guides/statistics/sampling/populations-samples-parameters-and-statistics

Today is all about “definitions”:
In most disciplines we learn “how to talk before walk"

Source: https://marketbusinessnews.com/financial-glossary/jargon-definition/

Some jargon is key to optimize communication & understand concepts more clearly

What is a population, a sample, an observation, a variable, a parameter?

Jargon gets an unfair treatment but learning and working in most fields would be very difficult without it.

Jargon is a real time saver!

Statistics are based on samples!

The most important goal of statistics is to **infer an unknown quantity** (e.g., height) of a population (e.g., plants of a particular species in Montreal) based on sample data!

infer an unknown quantity =
produce information about some chosen statistical **POPULATION** (e.g., number of trees, potation chip bags produced by a factory in a year) of interest!

Statistical (important) definitions: **POPULATION**

Entire collection of individual units (or observation units) that share a property or sets of properties from which you want to generalize knowledge about unknown quantities (observations) based on a sub-set of individual units (sample).

Examples -

- Coffee drinkers in Quebec.
- Coffee drinkers in Canada.
- Coffee drinkers in Canada that run in the morning.

Statistical (important) definitions: **POPULATION**

Entire collection of **individual units** that share a **property** or sets of properties from which you want to generalize knowledge about unknown quantities (**observations**) based on a sub-set of individual units (**sample**).

What is the average height and average weight of coffee drinkers in Canada that run in the morning?

Individual unit (or observation unit) = someone living in Canada that drinks coffee and runs in the morning.

Statistical (important) definitions: **POPULATION**

Entire collection of **individual units** that share a **property** or sets of properties from which you want to generalize knowledge about unknown quantities (**observations**) based on a sub-set of individual units (**sample**).

What is the average height and average weight of coffee drinkers in Canada that run in the morning?

Individual unit (or observation unit) = someone living in Canada that drinks coffee and runs in the morning.

Properties = Live in Canada, drink coffee and run in the morning.

Statistical (important) definitions: **POPULATION**

Entire collection of **individual units** that share a **property** or sets of properties from which you want to generalize knowledge about unknown quantities (**observations**) based on a sub-set of individual units (**sample**).

What is the average height and average weight of coffee drinkers in Canada that run in the morning?

Individual unit (or observation unit) = someone living in Canada that drinks coffee and runs in the morning.

Properties = Live in Canada, drink coffee and run in the morning.

Observation (or data point) = set of one or more quantities (measurements) on a single observation unit; ex. the weight and height of someone living in Canada that drinks coffee and run in the morning.

Statistical (important) definitions: **POPULATION**

Entire collection of **individual units** that share a **property** or sets of properties from which you want to generalize knowledge about unknown quantities (**observations**) based on a sub-set of individual units (**sample**).

What is the average height and average weight of coffee drinkers in Canada that run in the morning?

Individual unit (or observation unit) = someone living in Canada that drinks coffee and runs in the morning.

Properties = Live in Canada, drink coffee and run in the morning.

Observation (or data point) = set of one or more quantities (measurements) on a single observation unit; ex. the weight and height of someone living in Canada that drinks coffee and run in the morning.

Sample = subset of observation units form all possible observations in the population.

A **sample** of 11 individuals from the target **population** (**PROPERTIES:** Canadians that drink coffee and run in the morning).

Individual unit (or observation unit) = a Canadian that drinks coffee and runs in the morning.

Individual	Weight (kg)	Height (cm)
1	75.5	172
2	55.3	152
3	61.2	164
4	50.3	148
5	99.4	192
6	66.2	171
7	75.3	169
8	74.6	182
9	60.5	162
10	93.5	184
11	73.6	169

Observation (or data point) = set of one or more quantities (measurements) on a single observation unit; ex. the weight and height of someone living in Canada that drinks coffee and run in the morning.

TWO different observations are in red squares.

Statistical (important) definitions: **POPULATION**

Often the size of the population is unknown (i.e., we don't know how many people living in Canada drink coffee and run in the morning).

In many cases it can be so big that we consider the population as "infinite" for many purposes (more later on this during the semester).

Populations are also dynamic in the sense that they change through time.

ARTICLE

doi:10.1038/nature14967

Mapping tree density at a global scale

T. W. Crowther[1], H. B. Glick[1], K. R. Covey[1], C. Bettigole[1], D. S. Maynard[1], S. M. Thomas[2], J. R. Smith[1], G. Hintler[1], M. C. Duguid[1], G. Amatulli[3], M.-N. Tuanmu[3], W. Jetz[1,3,4], C. Salas[5], C. Stam[6], D. Piotto[7], R. Tavani[8], S. Green[9,10], G. Bruce[9], S. J. Williams[11], S. K. Wiser[12], M. O. Huber[13], G. M. Hengeveld[14], G.-J. Nabuurs[14], E. Tikhonova[15], P. Borchardt[16], C.-F. Li[17], L. W. Powrie[18], M. Fischer[19,20], A. Hemp[21], J. Homeier[22], P. Cho[23], A. C. Vibrans[24], P. M. Umunay[1], S. L. Piao[25], C. W. Rowe[1], M. S. Ashton[1], P. R. Crane[1] & M. A. Bradford[1]

A study led by Yale University researchers has found that there are over 3 trillion trees on Earth - but they are disappearing at an alarming rate.

The study found that there are around 3.04 trillion trees on Earth, or around 422 for each person on the planet.

The number is a huge increase on the previous global estimate, which was just over 400 billion trees worldwide.

The study was based on on-the-ground data about the number of trees in more than 400,000 plots of forest from all continents except Antarctica.

Source - https://www.independent.co.uk/environment/how-many-trees-are-there-on-earth-10483553.html

Let's think about this question:

If the number of trees is much bigger than we thought, does that mean that past estimates of the average size of trees on the planet are completely wrong?

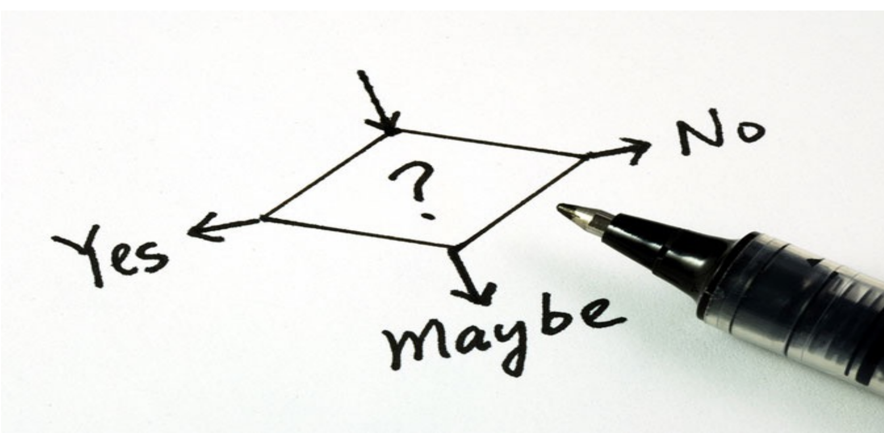
Statistical populations *versus* biological populations (let's not mix the two)

In biology, a **population** is all the organisms of the same group or species living in a particular geographical area.

In statistics, a **population** is a set of similar items (living or not) which is of interest to tackle a question of interest.

Let’s take a break - 2 minutes

Statistics is the science of assisting in decision making with incomplete knowledge based on populations that too often have unknown sizes (i.e., number of individual units)



Statistical (important) definitions: **POPULATION**

Examples -

Stars in the sky (“infinite”)
Sand in a river (“infinite”)
Countries in Europe (finite)
Bags of potato chips in a factory (finite)

To avoid confusion we often use the term “statistical population” instead of just “population”

Questions of interest influence the observational units and how the statistical population is defined

Question	Observational unit	Statistical population
What proportion of the plants are flowering?	An individual plant	All the plants in the ecological population
How many seeds per flower?	An individual plant in flower	All the plants in flower
How many seeds per white-flowered plant?	An individual white-flowered plant	All the white-flowered plants in flower
How many plants/m^2 in the field?	An area of m^2	All the areas of m^2 in the field
How long are the stamens?	A stamen	All the stamens
How much time do bees spend on a visit to a flower?	A visit by a bee to a flower	All the visits made by bees to flowers
How many bees visit in a 5-minute observation period?	A 5-minute observation period	All the 5-minute observation periods which could be made

- From David Heath, An Introduction To Experimental Design And Statistics For Biology

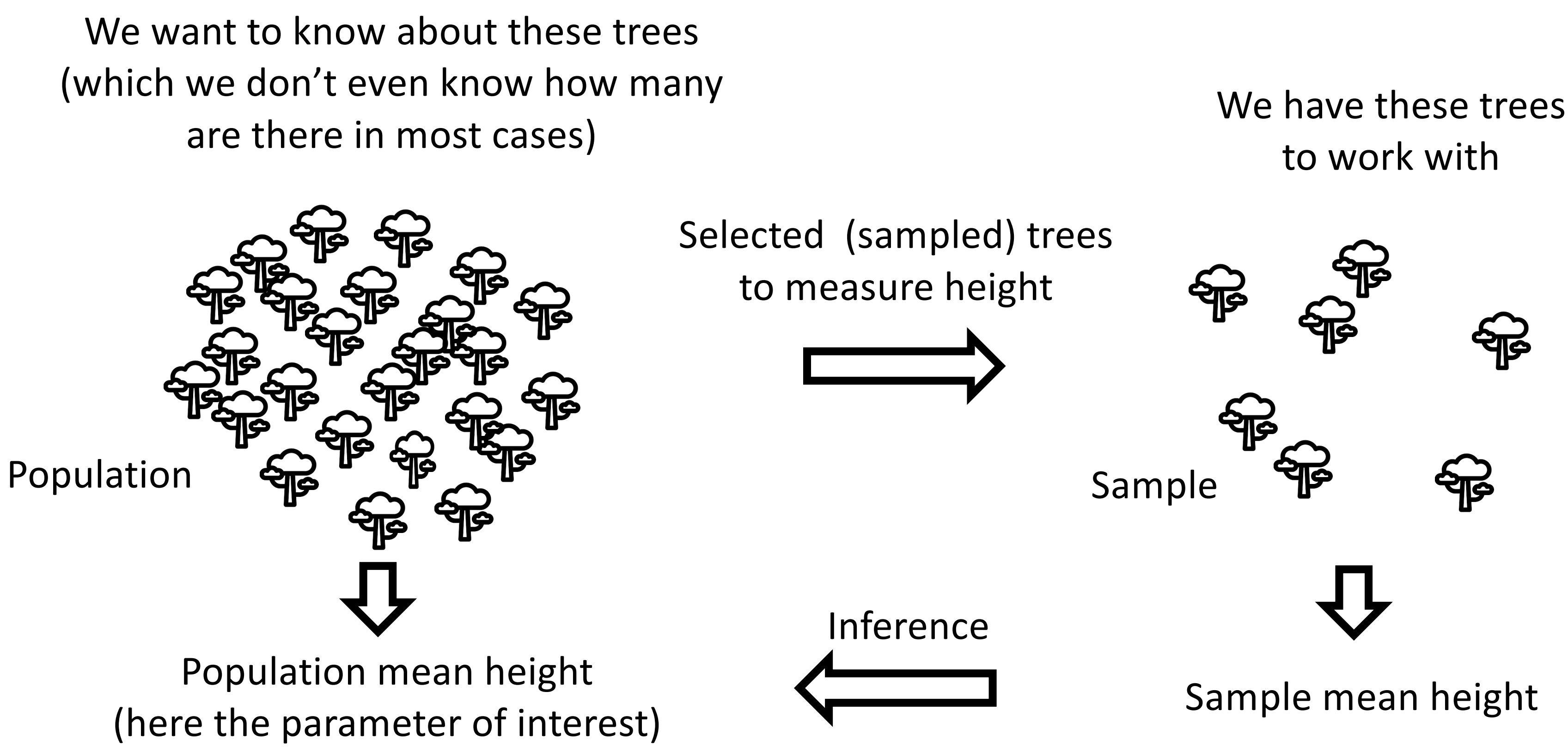
A **parameter** is a quantity describing a statistical population, whereas an **estimate** (or statistic) is the same quantity but calculated from a sample.

What is the average height of trees across all species in the world?

The parameter of interest (unknown) here is the average height all trees.

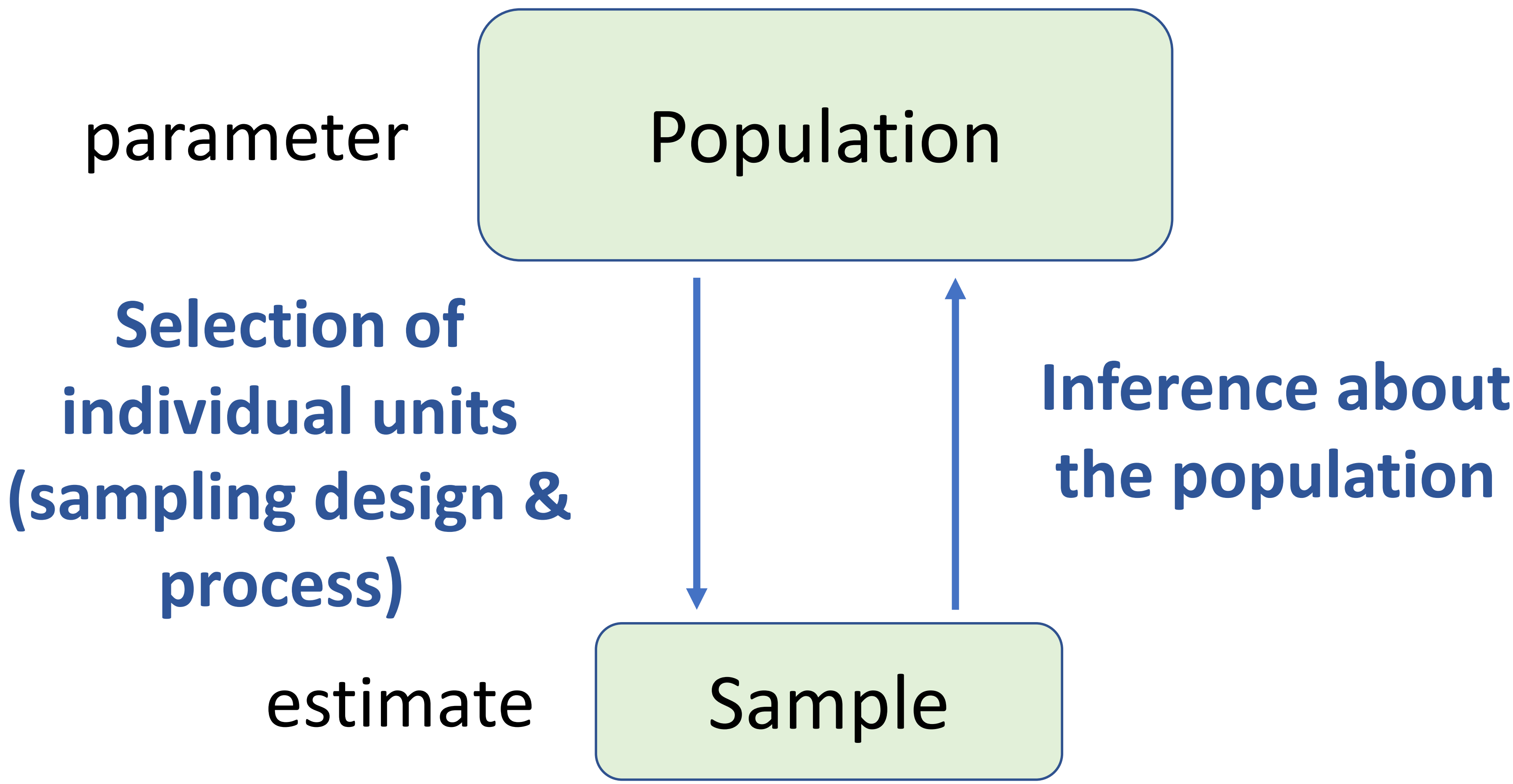
The estimate of interest (known) here are the average height of a smaller group of trees (sample).

The most important goal of statistics is to **infer an unknown quantity** (e.g., height) of a population based on sample data!



Inspired by https://www.cliffsnotes.com/study-guides/statistics/sampling/populations-samples-parameters-and-statistics

The most important goal of statistics is to **infer an unknown quantity** (e.g., height) of a population based on sample data!



Variables!

- A **variable** is any characteristic, number, or quantity that can be measured or counted and varies among observation units.

  Height, weight, age, gender, eye color, are examples of **variables**.

- It is key to be able to recognized the type of variable because they often guide the type of statistical analyses!

  - **Variables** (e.g., height, biomass) differ among **observation units** (e.g., one individual tree).

A **sample** of 11 individuals from the target populations (**PROPERTIES:** Canadians that drink coffee and run in the morning).

Individual unit (or observation unit) = a Canadian that drinks coffee and runs in the morning.

VARIABLES

Individual	Weight (kg)	Height (cm)
1	75.5	172
2	55.3	152
3	61.2	164
4	50.3	148
5	99.4	192
6	66.2	171
7	75.3	169
8	74.6	182
9	60.5	162
10	93.5	184
11	73.6	169

An **observation** contains all the values for the **variables** of interest such as the height and weight of a Canadian that drinks coffee and runs in the morning.

TWO different observations.

Types of variables

CATEGORICAL VARIABLES - describe membership in a category or group; characteristics of observations that do not have magnitude on a numerical scale. They can be:

Nominal (name)

- Survival (alive or dead),
- Method of disease transmission (e.g., water, air, animal vector),
- Eye colors (amber, blue, brown, gray, green, hazel, or red),
- Breed of a dog (e.g., collie, shepherd, terrier).

or Ordinal (ordered)

- Life stage (e.g., egg, larva, juvenile, adult),
- Snake bite severity score (e.g., minimal, moderate, severe),
- Size class (e.g., small, medium, large).

Types of variables

NUMERICAL VARIABLES - characteristics of observations have magnitude on a numerical scale.

Continuous (can take any real-number value)

- Core body temperature (e.g., degrees Celcius, ^{o}C),
- Territory size of a bird (e.g., hectares),
- Size of fish (e.g., cm)

Discrete (only take indivisible units)

- Age at death (e.g., years),
- Number of amino acids in a protein,
- Number of eggs in a bird nest.

Types of variables

NUMERICAL VARIABLES - characteristics of observations have magnitude on a numerical scale.

Continuous (can take any real-number value)

- Core body temperature (e.g., degrees Celcius, ^{o}C),
- Territory size of a bird (e.g., hectares),
- Size of fish (e.g., cm)

Discrete (only take indivisible units)

- Age at death (e.g., years),
- Number of amino acids in a protein,
- Number of eggs in a bird nest.

    - **How many numbers between 2 continuous numbers?**
    - **How many numbers between 2 discrete numbers?**

Statistical variables

Variables are not based on their measuring units (e.g., cm) but rather their types (e.g., height, length)

Arm length and leg length can be both measured in centimeters, but they are TWO different variables.

Enjoy your definitions!

Source: https://marketbusinessnews.com/financial-glossary/jargon-definition/