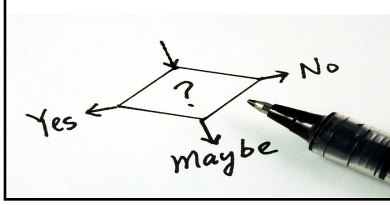
Don't hesitate to raise your hand during lectures if you have any questions.

I also try to "read the room" and will step in when I sense that students may be unsure or have questions they haven't yet voiced.

1

Statistics in the science of supporting decision-making under incomplete information / knowledge



2

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

3

ARTICLE

doi:10.1038/nature14967

Mapping tree density at a global scale

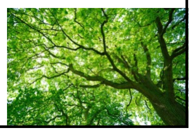
T. W. Crowther[1], H. B. Glick[1], K. R. Covey[1], C. Bettigole[1], D. S. Maynard[1], S. M. Thomas[2], J. R. Smith[1], G. Hintler[1], M. C. Duguid[1], G. Amatulli[3], M.-N. Tuanmu[3], W. Jetz[1,3,4], C. Salas[5], C. Stam[6], D. Piotto[7], R. Tavani[8], S. Green[9,10], G. Bruce[9], S. J. Williams[11], 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

4

Statistics are based on samples!

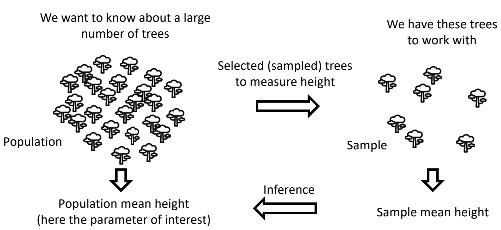
A central goal of statistics is to learn about an entire population (e.g., the average height of a plant species) using information from a sample - a smaller subset of observations or individuals drawn from that population.

Biologists are relatively small

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

5

A central goal of statistics is inference: drawing conclusions about unknown population quantities (e.g., average height of a tree species in a forest) from **sample** data.



A sample is a subset of observations or individuals drawn from a larger population, used to learn about that population.

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

6

Today focuses on definitions!!

In most disciplines, learning the language comes before learning the methods.

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

7

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!

8

Statistical inference is based on samples!

The central goal of statistics is inference: drawing conclusions about unknown population quantities (e.g., average height of a tree species in a forest) from sample data.

Infer an unknown quantity =
use a sample to produce information (e.g., an average value) about a chosen statistical population of interest (e.g., trees in a forest; potato-chip bags produced by a factory in a year).

9

Statistical (important) definitions: **POPULATION**

A population is the entire collection of individual units (or observation units) that share a common property or set of properties.

It is from this group that we aim to generalize knowledge about unknown quantities (e.g., mean of the population) based on a subset of individual units, known as a sample.

Examples -

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

10

Statistical (important) definitions: **POPULATION**

A population is the entire collection of **individual units** (or **observation** units) that share a common property or set of properties. It is from this group that we aim to generalize knowledge about unknown quantities based on a subset of individual units (observation), known as a **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.

11

Statistical (important) definitions: **POPULATION**

A population is the entire collection of **individual units** (or **observation** units) that share a common property or set of properties. It is from this group that we aim to generalize knowledge about unknown quantities based on a subset of individual units (observation), known as a **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.

12

Statistical (important) definitions: **POPULATION**

A population is the entire collection of **individual units** (or **observation** units) that share a common property or set of properties. It is from this group that we aim to generalize knowledge about unknown quantities based on a subset of individual units (observation), known as a **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.

13

Statistical (important) definitions: **POPULATION**

A population is the entire collection of **individual units** (or **observation** units) that share a common property or set of properties. It is from this group that we aim to generalize knowledge about unknown quantities based on a subset of individual units (observation), known as a **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 from all possible observations in the population.

14

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

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.

15

Statistical (important) definitions: **POPULATION**

The size of a population is often unknown. For example, we may not know how many people in Canada drink coffee and go for a run in the morning.

In many cases, the population can be so large that it is considered 'infinite' for practical purposes (we will explore this concept further later in the semester).

Additionally, populations are dynamic, meaning they can change over time.

ARTICLE

doi:10.1038/nature14967

Mapping tree density at a global scale

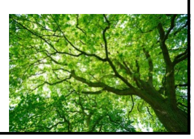
T. W. Crowther[1], H. B. Glick[1], K. R. Covey[1], C. Bettigole[1], D. S. Maynard[1], S. M. Thomas[2], J. R. Smith[1], G. Hintler[1], M. C. Duguid[1], G. Amatulli[3], M.-N. Tuanmu[3], W. Jetz[1,3,4], C. Salas[5], C. Stam[6], D. Piotto[7], R. Tavani[8], S. Green[9,10], G. Bruce[9], S. J. Williams[11], 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.

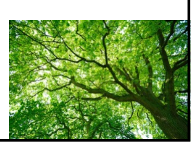
The study was based on on-the-ground data about the number of trees in more than 400,000 plots of forest from all continents except Antarctica.



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

Let's pause and think:

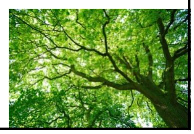
If the number of trees is revised upward (from 400 billion to 3 trillion), should we automatically conclude that earlier estimates of the average tree size were wrong?



Let's pause and think:

If the number of trees is revised upward (from 400 billion to 3 trillion), should we automatically conclude that earlier estimates of the average tree size were wrong?

Not necessarily. A higher population size changes **how many trees exist**, but it doesn't automatically change **the average size** - that depends on *which trees* are being counted and how the average was estimated.



19

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

In biology, a population refers to all organisms of the same group or species that live in a specific geographical area.

In statistics, a population is a set of similar items, whether living or non-living, that are of interest for answering a particular research question.

For clarity, it's best to refer to it as a 'statistical population' when the goal is to infer quantities from it.

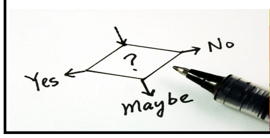
20

Let's take a break – 1 minute

21

Statistics is the science of supporting decision-making under incomplete information, most often by analyzing samples drawn from populations of unknown size.

In other words: Statistics is the science of inference and decision-making under uncertainty, using samples to learn about populations we cannot fully observe.




22

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)

Again, to avoid confusion, we often use the term “statistical population” instead of just “population”; and some refer to “statistical universe”.

23

What we want to know determines what we measure and which statistical population we are studying.

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 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

24

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

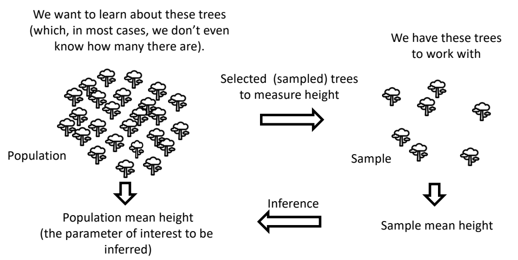
What is the average height of trees across all species worldwide?

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

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

25

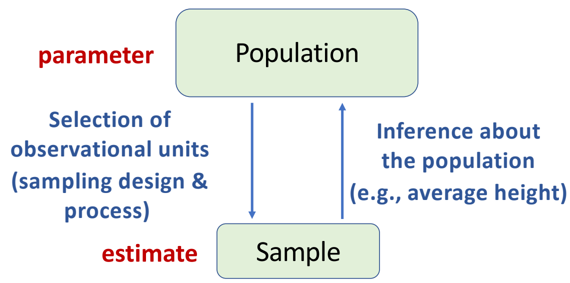
A central goal of statistics is inference: drawing conclusions about unknown population quantities (e.g., average height of a tree species in a forest) from **sample** data.



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

26

A central goal of statistics is inference: drawing conclusions about unknown population quantities (e.g., average height of a tree species in a forest) from **sample** data.



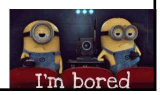
27

Variables!

A *variable* is any characteristic, number, or quantity that can be measured or counted and varies among observation units. Examples of variables include height, weight, age, gender, and eye color.

Recognizing the type of *variable* is crucial, as it often determines the appropriate type of statistical analysis.

Variables (e.g., height, biomass) differ among observation units (e.g., individual trees).



28

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 unit*** (or simply observation) contains the values of all ***variables*** of interest, such as the height and weight of a Canadian who drinks coffee and runs in the morning.

TWO different observations.

29

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).

30

Types of variables

The case of YEARS, MONTHS and WEEKDAYS

Although these categories have a natural order (e.g., January comes before February; Monday comes before Tuesday), they are typically considered nominal because the sequence does not imply any inherent numerical value or distance between the categories.

In some contexts, they can be treated as ordinal if the order is important (e.g., time series analysis), but generally, they are considered nominal variables.

31

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, ºC),
- 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.

32

Important knowledge in mathematics and statistics:

How many numbers lie between 2 continuous numbers?

How many numbers lie between 2 discrete numbers?

33

Statistical variables

Variables are defined by their type (e.g., height, length), not by their measurement units (e.g., cm).

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

34

Enjoy your definitions!

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

35