1.1 Statistics Terminology

Dr. Lauren Perry

Goals

1. Learn basic statistical terminology.
   - Understand key terms*
   - Read a data matrix
   - Identify variable types

*key terms are shown in bold in both the slides and the course notes.

There are two ways to think about statistics:

1. **Descriptive statistics** are methods for *describing* information.
   - Ex: 66% of eligible voters voted in the 2020 presidential election (the highest turnout since 1900!).
2. **Inferential statistics** are methods for *drawing inference* (making decisions about something we are uncertain about).
   - Ex: A poll suggests that 75% of voters will select a Candidate A. We reasonably conclude that Candidate A will win the election.

Data is factual information. We collect data from a **population**, the collection of all individuals or items a researcher is interested in.

- Collecting data from an entire population is called a **census**.
  - This is complicated and expensive!
- We can also take a **sample**, a subset of the population we get data from.

Checkpoint 1

Identify the population and the sample.

A. A survey of 2084 US households found that 45% have multiple TVs.

B. A local university wants to impose a new student fee in order to offer a better student rec center. They ask 87 students whether they support this fee.

C. A scientist wants to track the life cycles of invasive Burmese pythons in Florida. She spends a month in the field and tags 52 pythons for monitoring.

Data are often organized in what we call a **data matrix**. If you've ever seen data in a spreadsheet, that's a data matrix!

	Age	Gender	Smoker	Marital Status
Person 1	45	Male	yes	married
Person 2	23	Female	no	single
Person 3	36	Other	no	married
Person 4	29	Female	no	single

- Each row represents one **observation** (also called **cases** or **subjects**). These are the individuals or items in the sample.
- Each column represents a **variable**, the characteristic or thing being measured.

Checkpoint 2

The following table shows part of the data matrix from a Stat 1 course survey.

	Age	Year	Major	Current Units
1	19	Sophomore	Health Sciences	15
2	19	Sophomore	Business	15
3	19	Sophomore	Undecided	14
⋮	⋮	⋮	⋮	⋮
29	21	Junior	Business	15

1. What does each row of the data matrix represent?
2. What does each column of the data matrix represent?

Variable Types

There are two types of variable:

Numeric or **quantitative** variables take *numeric* values AND it is sensible to do math with those values.

Categorical or **qualitative** variables take values that are *categories*.

Variable Types: Numeric

We can further break down numeric variables into

- **Discrete numeric** variables take numeric values with jumps. Typically, this means they can only take whole number values. These are often counts of something.
  - Ex: the number of pets you have, the number of cars that drive through an intersection during rush hour, or the number of classes students are taking
- **Continuous numeric** variables take values "between the jumps". Typically, this means they can take decimal values.
  - Ex: weights of guinea pigs, milliliters of medication administered, or any measurements of time

Variable Types: Categorical

We can break categorical variables down into

- **Ordinal categorical** variables have categories with some kind of intrinsic ordering, meaning we can rank them in a meaningful way.
  - Ex: a survey asking for approval levels might have categories "strongly disapprove, disapprove, neutral, approve, strongly approve"; and letter grades have the standard ordering "A, B, C, D, F"
- **Nominal categorical** variables have categories with no intrinsic ordering.
  - Ex: eye color, high school attended, and the city people live in

The "Does it make sense"? Test

- If you're unsure whether a variable is discrete or continuous, pick a number with some decimal place and ask yourself if that value makes sense. If it doesn't, it's probably discrete.
- Sometimes, categories can be represented by numbers. Ask yourself if it makes sense to do math with those numbers. If it doesn't make sense, it's probably a categorical variable.

Checkpoint: Determine the variable type.

A. species

B. temperature in Celsius

C. level of education

D. blood type

E. grams of flour in a cake recipe

F. political party

G. level to which a person agrees with some statement

H. number of siblings

I. number of cars that cross a bridge during rush hour

J. heart rate (beats per minute)