Welcome to:

AP Statistics

Mr. Cedarlund

True or false

About AP Stats

Fastest growing of all 37 AP subjects

108,000

IB Math

1. Exploring data
2. Planning a study
3. Anticipating patterns
4. Statistical inference ✓

AP Stats

Overview

LEARNING TARGETS

By the end of this section, you should be able to:

- ✓IDENTIFY the individuals and variables in a set of data.
- ✓CLASSIFY variables as categorical or quantitative.

to "note" or
"not to note"

Organizing Data

Statistics is the science and art of *collecting*, *analyzing*, and *drawing conclusions* from data.

Organizing Data

Statistics is the science and art of *collecting*, *analyzing*, and *drawing conclusions* from data.

An **individual** is an object described in a set of data. Individuals can be people, animals, or things.

Organizing Data

Statistics is the science and art of *collecting*, *analyzing*, and *drawing conclusions* from data.

An **individual** is an object described in a set of data. Individuals can be people, animals, or things.

Variable - an attribute that can take different values for different individuals

Variable - an attribute that can take different values for different individuals

Categorical Variable
assigns labels that place each individual into a particular group, called a category.

Quantitative Variable
takes number values that are quantities—counts or measurements.

Categorical Variable
assigns labels that place each individual into a particular group, called a category.

Quantitative Variable
takes number values that are quantities—counts or measurements.

Values are usually words (but not always)

zipcodes 94563 37214

area codes 541 503

category.

Values are usually words (but not always)

zipcodes 94563 37214

area codes 541 503

quantitative variables like age, weight, etc can be divided into categories

NO carbs

-4

TON carbs

-2

it depends on how you want to use the variable

Analyzing Data

A **variable** generally takes values that vary. We are interested in the pattern of that variation.

Distribution

✓ tells us what values a variable takes and how often it takes those values.

Analyzing Data

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Analyzing Data

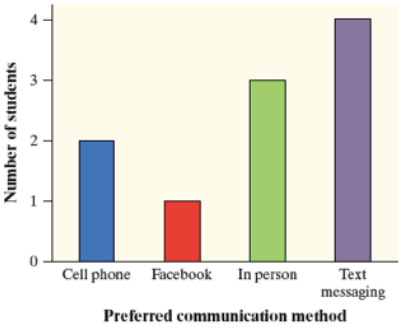
Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Categorical variable *"Preferred communication method"*

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Categorical variable *"Preferred communication method"*

Bar graph of *"Preferred communication method"*



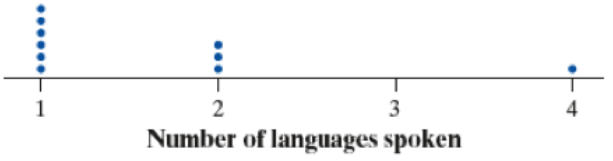
Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Quantitative variable *"Number of languages spoken"*

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Quantitative variable *"Number of languages spoken"*

Dot plot of *"Number of languages spoken"*



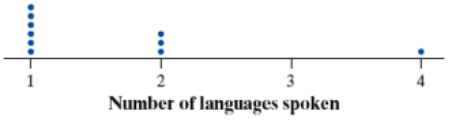
Examine each variable by itself. Then study relationships among the variables.

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Examine each variable by itself. Then study relationships among the variables.

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Start with graphs

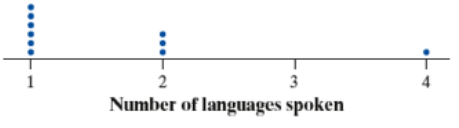


Number of languages spoken

Examine each variable by itself. Then study relationships among the variables.

Province	Gender	Number of languages spoken	Handedness	Height (cm)	Wrist circum-ference (mm)	Preferred communication
Saskatchewan	Male	1	Right	175.0	180	In person
Ontario	Female	1	Right	162.5	160	In person
Alberta	Male	1	Right	178.0	174	Facebook
Ontario	Male	2	Right	169.0	160	Cell phone
Ontario	Female	2	Right	166.0	65	In person
Nunavut	Male	1	Right	168.5	160	Text messaging
Ontario	Female	1	Right	166.0	165	Cell phone
Ontario	Male	4	Left	157.5	147	Text messaging
Ontario	Female	2	Right	150.5	187	Text messaging
Ontario	Female	1	Right	171.0	180	Text messaging

Start with graphs



Number of languages spoken

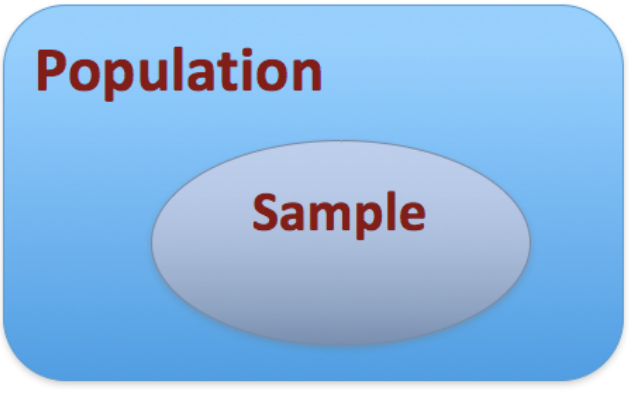
Add numerical summaries

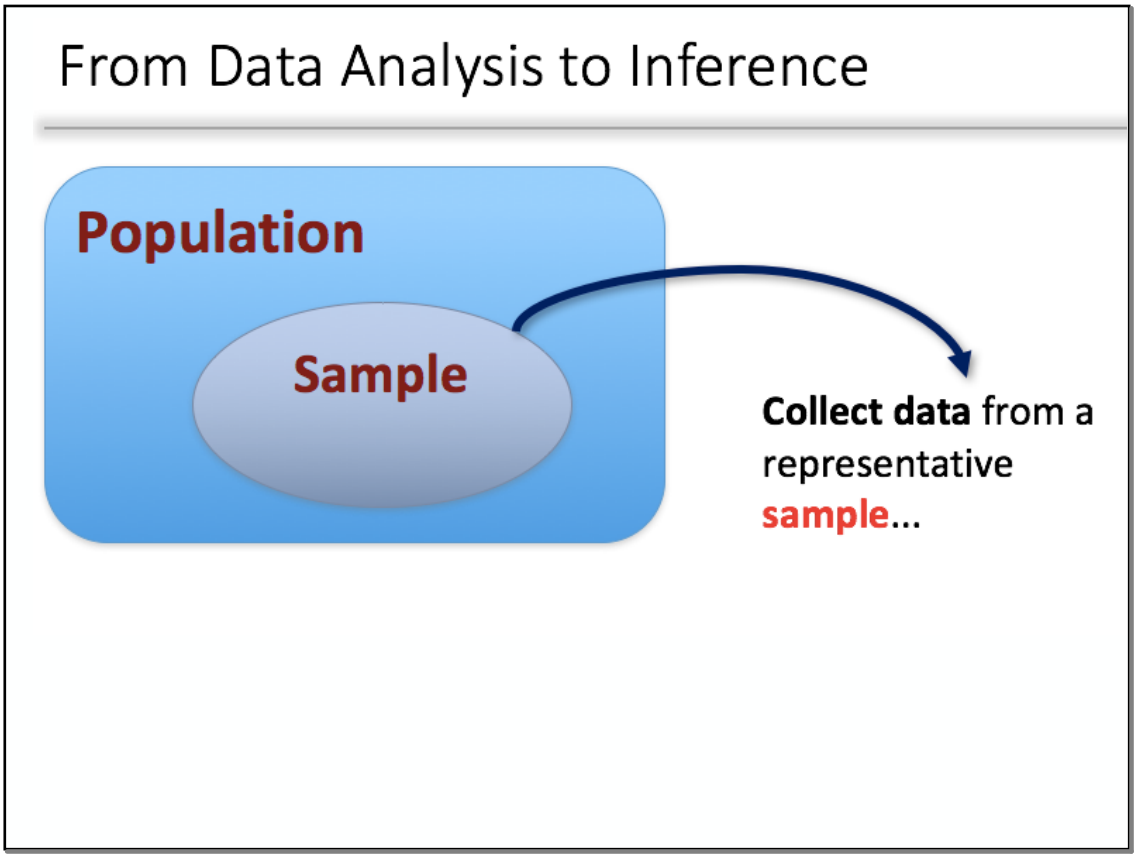
Descriptive Statistics: Number of languages spoken

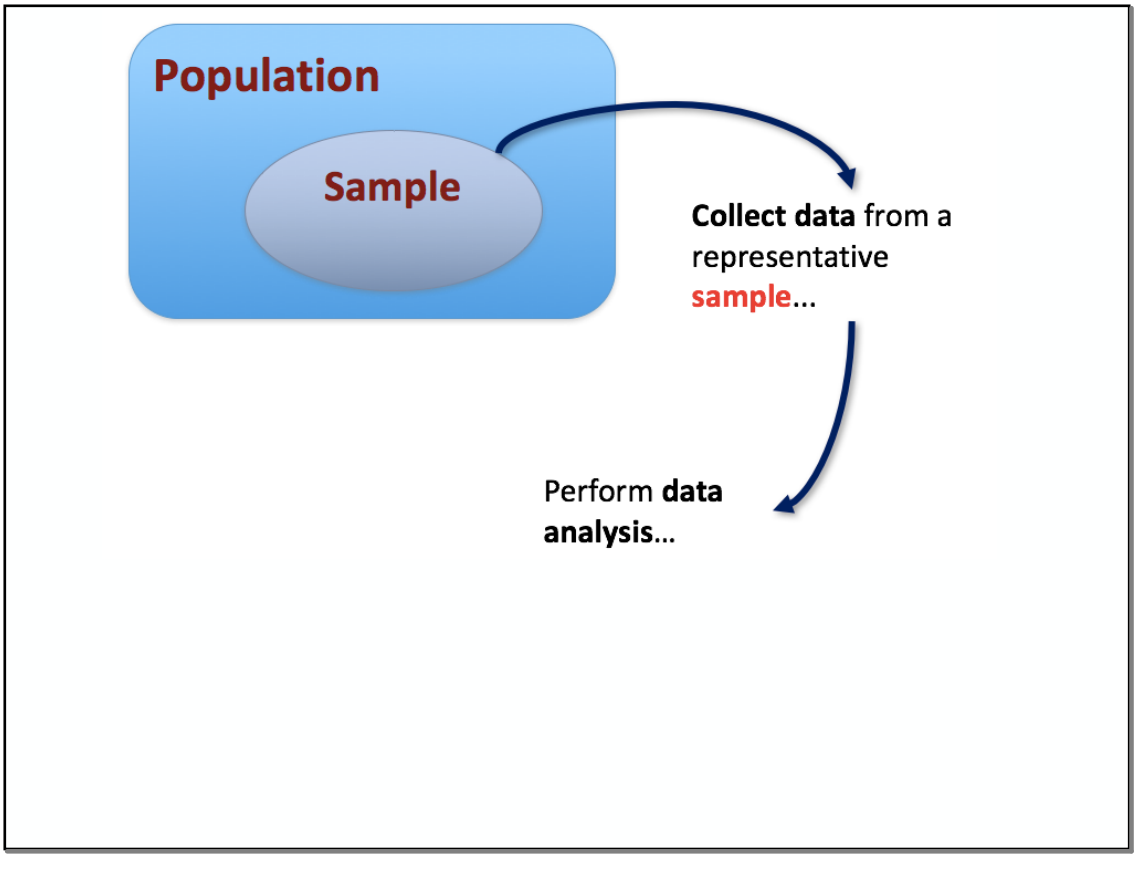
Variable	Total Count	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Number of languages spoken	10	1.600	0.966	1.000	1.000	1.000	2.000	4.000

From Data Analysis to Inference

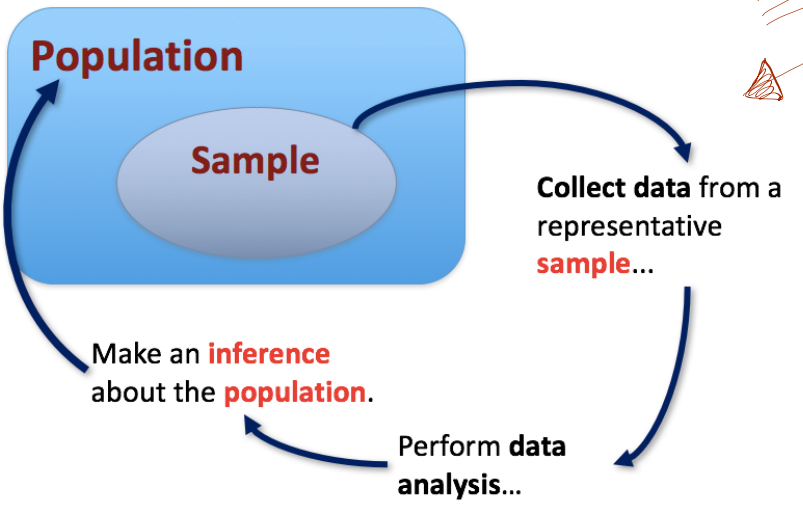
From Data Analysis to Inference



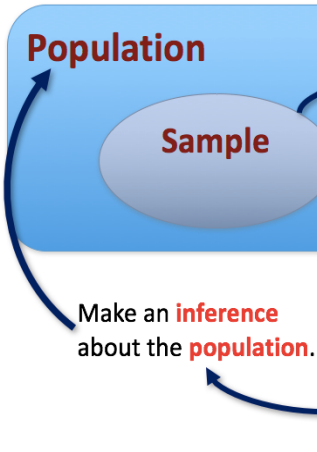

From Data Analysis to Inference
Population
Sample
Collect data from a representative sample...


Population
Sample
Collect data from a representative sample...
Perform data analysis...

From Data Analysis to Inference



Notes



what if a lady walks into a.....

SMELLING PARKINSON'S DISEASE

1. Why would it be important to know that someone can smell Parkinson's disease?

✓ early diagnosis
✓ avoid misdiagnosis

2. How many correct decisions (out of 12) would you expect Joy make if she couldn't really smell Parkinson's and was just guessing?

about ½

3. How many correct decisions (out of 12) would it take to *convince* you that Joy really could smell Parkinson's?

12 ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓
11 ✓ ✓
10 ✓

SIMULATING THE EXPERIMENT

Although the researchers wanted to believe Joy, there was a chance that she may not really be able to tell Parkinson's by smell. It's logical to be skeptical of claims that are very different than our experiences. If Joy couldn't really distinguish Parkinson's by smell, then she would just have been guessing which shirt was which. The researchers were not willing to commit time and resources to a larger investigation unless they could be convinced to that Joy's wasn't just guessing. When researchers have a claim that they suspect (or hope) to find evidence against, it's called the **null hypothesis**.

4. What claim were the researchers hoping to find evidence *against*? That is, what was their prior belief (**null hypothesis**) about the ability to smell Parkinson's?

H_0: Joy cannot smell Parkinson's

5. What claim were the researchers hoping to find evidence *for*? This is called the **alternative hypothesis** or the **research hypothesis**.

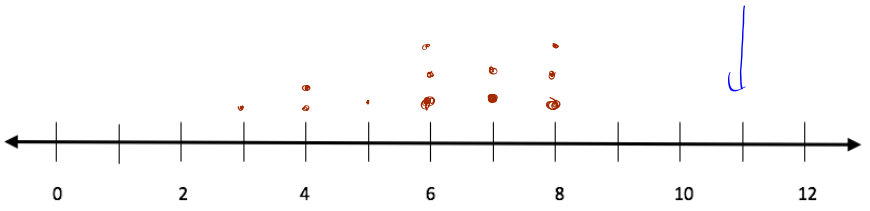
H_A H_1: Joy can smell Parkinson's

To investigate the idea that Joy was just guessing which shirt was worn by which type of person, we will assume that the null hypothesis is true.

6. Your instructor will hand you 12 cards (shirts) that have been shuffled into a random order. Don't turn them over yet! On the back of some of them is "Parkinson's" and on the back of others is "No Parkinson's." For each card, guess Parkinson's or No Parkinson's. Once you have made your guess, turn the card over and see if you were correct. Repeat this for each card and record the number of correct identifications (out of 12) below.

Tally of correct identifications	Number of correct identifications	Proportion of correct identifications

7. Create a dotplot of the number of correct identifications with the rest of the class. Record the results below.



8. In the actual experiment, Joy identified 11 of the 12 shirts correctly. Based on the very small-scale simulation by you and your classmates, what proportion of the simulations resulted in 11 or more shirts correctly identified, assuming that the person was guessing?

9. The proportion you just calculated is a crude estimate of a true probability called a ***P*-value**. How might we improve our estimate of the true probability?

STATISTICAL INFERENCE FROM THE SIMULATION

10. Use the One Categorical Variable, Single Group applet at stapplet.com to run this simulation 10000 times. Then use that simulation to get a (likely) better estimate of the *p*-value for 11 or more shirts correctly identified, assuming that this person was just guessing. Is it *possible* that Joy correctly identified 11 shirts just by random chance (guessing)? Is it *likely*?

11. An interesting side note is that Joy's one "mistake" really wasn't a mistake. The shirt was worn by a person who supposedly didn't have Parkinson's even though Joy claimed that she could smell the telltale smell on that shirt. That person called the experimenters a little while after the experiment and reported that he had just been diagnosed with Parkinson's disease. That meant that Joy correctly identified 12 out of 12 shirts. What is the approximate *P*-value for 12 shirts correctly identified, assuming that this person was just guessing?

Note: A small *P*-value is considered strong evidence against the null hypothesis and in favor of the alternative hypothesis. But how small is small? As a rule of thumb, statisticians generally agree that *P*-values below 0.05 provide pretty strong evidence against the null hypothesis. Observed results with small *P*-values are said to be **statistically significant**.

Assignment:

1. Getting to know your textbook

2. Read p. 2-5

   1.1 1, 5, 7, 9, 10

read syllabus