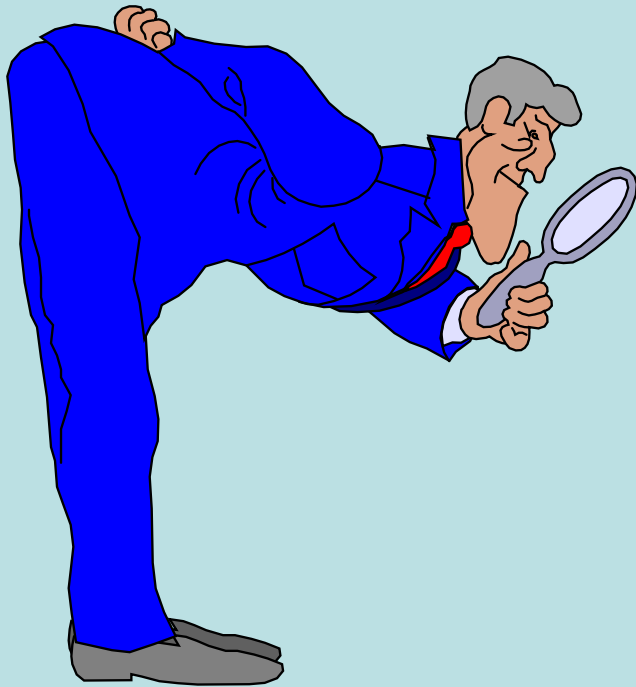


Statistical Inference

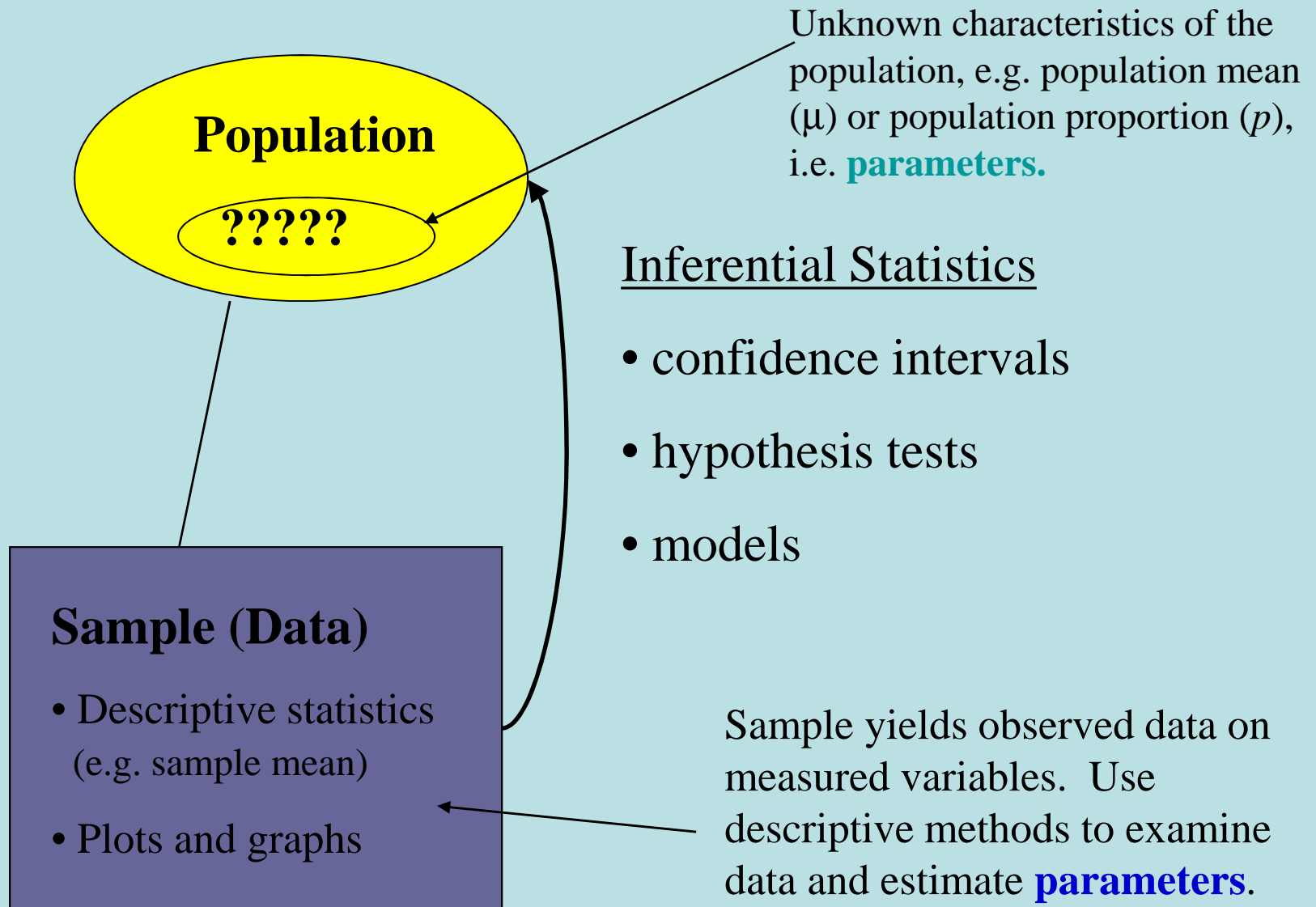
An introduction

Big picture



Use a random sample to learn something about a larger population.

Inference for a Single Population



Two main ways to learn about a population

- Confidence intervals
- Hypothesis testing

Confidence Intervals

- Allow us to use sample data to **estimate** a population value, like the true mean or the true proportion, i.e. estimate parameters.
- *Example:* What is the current mean GPA of U.S. college & university students ?

Hypothesis Testing

- Allows us to use sample data to **test a claim** about a population, such as testing whether a population proportion or population mean equals some number.
- *Example:* The mean GPA of U.S. college & university students today is larger than 2.70 which was the mean GPA in 1990?

General Idea of Hypothesis Testing

- Make an initial assumption.
- Collect evidence (data).
- Based on the available evidence, decide whether or not the initial assumption is reasonable.

Hmm? Let's illustrate this idea



Example: Grade inflation?

(Has mean GPA increased since 1990?)

Population of
5 million college
students



Sample of
100 college students

Is the average
GPA 2.7 ?

(Imagine that 2.7 was
mean GPA for U.S. college
students in 1990)

How likely is it that
100 students would
have an average
GPA as large as 2.9
**if the population
average was 2.7?**

Making the Decision

- It is either *likely* or *unlikely* that we would collect the evidence we did given the initial assumption.
- (*Note*: “Likely” or “unlikely” is measured by calculating a probability!)
- If it is *likely*, then we “**do not reject**” our initial assumption. There is not enough evidence to do otherwise.

Making the Decision (cont'd)

- If it is *unlikely*, then:
 - either our initial assumption is correct and we experienced an unusual event
 - or our initial assumption is incorrect
- In statistics, if it is unlikely, we decide to “**reject**” our initial assumption.

Idea of Hypothesis Testing: Criminal Trial Analogy

- First, state 2 hypotheses, the null hypothesis (“ H_0 ”) and the alternative hypothesis (“ H_A ”)
 - H_0 : Defendant is not guilty.
 - H_A : Defendant is guilty.

An aside:

Identification of hypotheses

- The **null hypothesis** always represents the status quo, *i.e.* the hypothesis that requires no change in current behavior.
- The **alternative hypothesis** is the conclusion that the researcher is trying to make.

Criminal Trial Analogy (continued)

- Then, collect evidence, such as finger prints, blood spots, hair samples, carpet fibers, shoe prints, ransom notes, handwriting samples, etc.
- In statistics, the **data** are the evidence.

Criminal Trial Analogy (continued)

- Then, make initial assumption.
 - Defendant is innocent until proven guilty.
- In statistics, we always **assume the null hypothesis is true.**

Criminal Trial Analogy (continued)

- Then, make a decision based on the available evidence.
 - If there is sufficient evidence (“beyond a reasonable doubt”), **reject the null hypothesis.** (Behave as if defendant is guilty.)
 - If there is not enough evidence, **do not reject the null hypothesis.** (Behave as if defendant is not guilty.)

Important Point About Errors

- Neither decision entails proving the null hypothesis or the alternative hypothesis.
- We merely state there is enough evidence to behave one way or the other.
- This is also always true in statistics! No matter what decision we make, there is **always a chance we made an error.**
- Because there are two possible decisions there are two potential errors that can be made, **Type I and II Errors.**
(see separate ppt for more about these)