Introduction

Chapter 1 of Social Statistics

Chris Lawrence

cnlawren@olemiss.edu

Introduction

In this chapter, we will discuss:

- What statistics are

Introduction

In this chapter, we will discuss:

- What statistics are
- The difference between a statistic and a parameter

Introduction

In this chapter, we will discuss:

- What statistics are
- The difference between a statistic and a parameter
- Samples and populations

Introduction

In this chapter, we will discuss:

- What statistics are
- The difference between a statistic and a parameter
- Samples and populations
- Descriptive and inferential statistics

Why care about statistics?

Understanding statistics is *important*.
If you don't understand statistics, then whenever you come across statistically-based information, you'll have just two choices:

- Unconditionally accept the information.

Why care about statistics?

Understanding statistics is *important*.
If you don't understand statistics, then whenever you come across statistically-based information, you'll have just two choices:

- Unconditionally accept the information.
- Find someone else to interpret it for you.

We are surrounded by statistics

We encounter statistics all of the time in everyday life, even though their role is not always obvious!

- Car and health insurance rates are based on stats.

We are surrounded by statistics

We encounter statistics all of the time in everyday life, even though their role is not always obvious!

- Car and health insurance rates are based on stats.
- Politicians make decisions based on public opinion polls.

We are surrounded by statistics

We encounter statistics all of the time in everyday life, even though their role is not always obvious!

- Car and health insurance rates are based on stats.
- Politicians make decisions based on public opinion polls.
- We rely on stats for crime rates, consumer information, and weather reports.

What is political science?

For a moment, let us revisit Shively. . .

- The goal of social science is to describe and explain different types of social phenomena.

What is political science?

For a moment, let us revisit Shively. . .

- The goal of social science is to describe and explain different types of social phenomena.
- In political science, we want to describe and explain *political* phenomena.

Characteristics and variables

- So, we want to understand how social or political characteristics relate to and affect one another.

Characteristics and variables

- So, we want to understand how social or political characteristics relate to and affect one another.
- These *characteristics* are also called *variables*.

All variables must vary

- Variables **vary**. Things that don’t vary can’t be variables.

All variables must vary

- Variables **vary**. Things that don't vary can't be variables.
- Variables change in some measurable way from unit to unit.

All variables must vary

- Variables **vary**. Things that don't vary can't be variables.
- Variables change in some measurable way from unit to unit.
- For example, a person's weight, income, and party identification are examples of variables.

Dependent and independent variables

- When we examine relationships between variables, we normally examine the change in one variable that is caused by another.

Dependent and independent variables

- When we examine relationships between variables, we normally examine the change in one variable that is caused by another.
- The variable that is being changed is the **dependent** variable.

Dependent and independent variables

- When we examine relationships between variables, we normally examine the change in one variable that is caused by another.
- The variable that is being changed is the **dependent** variable.
- The variable or variables that cause change are the **independent** variable or variables.

Characteristics, attitudes, and behavio

There are three major types of variable we can measure:

- Characteristics: gender (male/female), religion.

Characteristics, attitudes, and behavio

There are three major types of variable we can measure:

- Characteristics: gender (male/female), religion.
- Attitudes: like/dislike math.

Characteristics, attitudes, and behavio

There are three major types of variable we can measure:

- Characteristics: gender (male/female), religion.
- Attitudes: like/dislike math.
- Behavior: embrace/avoid math.

Examples

- American politics: Republican, hates Al Gore, voted for George W. Bush.

Examples

- American politics: Republican, hates Al Gore, voted for George W. Bush.
- German politics: Social Democrat, likes European Union, voted for SPD.

Examples

- American politics: Republican, hates Al Gore, voted for George W. Bush.
- German politics: Social Democrat, likes European Union, voted for SPD.
- International relations: Country X is a theocracy, intolerant of other religions, goes to war with infidels

Populations versus samples

A *population* is everyone or everything that is in our universe of interest.

- Examples: everyone in the United States, all of the countries in the world, every fraternity member at Ole Miss, every pop singer between 15 and 30.

Populations versus samples

A *population* is everyone or everything that is in our universe of interest.

- Examples: everyone in the United States, all of the countries in the world, every fraternity member at Ole Miss, every pop singer between 15 and 30.
- The population we examine is usually determined by our research question.

Populations versus samples

A *population* is everyone or everything that is in our universe of interest.

- Examples: everyone in the United States, all of the countries in the world, every fraternity member at Ole Miss, every pop singer between 15 and 30.
- The population we examine is usually determined by our research question.
- For example, if we wanted to know why people vote the way they do, our population would be voters.

Populations can be big

So if we care about vote choice, we would have a very large population:

- Approximately 270 million citizens in the U.S.

Populations can be big

So if we care about vote choice, we would have a very large population:

- Approximately 270 million citizens in the U.S.
- Approximately 180 million are over 18 (2/3)

Populations can be big

So if we care about vote choice, we would have a very large population:

- Approximately 270 million citizens in the U.S.
- Approximately 180 million are over 18 (2/3)
- Approximately 90 million of those vote (1/2)

Parameters: math about populations

When we have some math that tells us something about an entire population, we call that math a *parameter*.

- The average income of all Americans.

Parameters: math about populations

When we have some math that tells us something about an entire population, we call that math a *parameter*.

- The average income of all Americans.
- The percentage of all voters who are Democrats.

Parameters: math about populations

When we have some math that tells us something about an entire population, we call that math a *parameter*.

- The average income of all Americans.
- The percentage of all voters who are Democrats.
- The average GPA of all fraternity members at Ole Miss.

Parameters: math about populations

When we have some math that tells us something about an entire population, we call that math a *parameter*.

- The average income of all Americans.
- The percentage of all voters who are Democrats.
- The average GPA of all fraternity members at Ole Miss.
- The number of countries in the world that have a theocratic government.

Parameters: math about populations

When we have some math that tells us something about an entire population, we call that math a *parameter*.

- The average income of all Americans.
- The percentage of all voters who are Democrats.
- The average GPA of all fraternity members at Ole Miss.
- The number of countries in the world that have a theocratic government.
- The percentage of pop stars from 15 to 30 who have had cosmetic surgery.

Parameters: about the population

- Parameters are always *descriptive*: they always describe everyone in the population.

Parameters: about the population

- Parameters are always *descriptive*: they always describe everyone in the population.
- When we talk about parameters, we use lower case greek letters, like σ (sigma). The size of the population is N.

Parameters: about the population

- Parameters are always *descriptive*: they always describe everyone in the population.
- When we talk about parameters, we use lower case greek letters, like σ (sigma). The size of the population is N.
- Parameters are usually collected through a procedure like a census or by keeping track of every event that happens.

Samples

- When we don’t have the entire population, we have a *sample*.

Samples

- When we don’t have the entire population, we have a *sample*.
- For example, we can’t collect information on all voters (why not?), so we have to take a sample.

Samples

- When we don’t have the entire population, we have a *sample*.
- For example, we can’t collect information on all voters (why not?), so we have to take a sample.
- With an appropriate sample, we can estimate the parameters of the population.

Sampling from a population

- To estimate the parameters, we need a *representative* sample that is *randomly* selected.

Sampling from a population

- To estimate the parameters, we need a *representative* sample that is *randomly* selected.
- For example, we can estimate the percentage of voters who are Democrats by getting data from a sample of voters.

Statistics: math based on samples

- **Statistics** are the mathematics associated with samples.

Statistics: math based on samples

- **Statistics** are the mathematics associated with samples.
- We usually use upper-case Latin letters for statistics (like S). The size of a sample is denoted by n.

Statistics: math based on samples

- **Statistics** are the mathematics associated with samples.
- We usually use upper-case Latin letters for statistics (like S). The size of a sample is denoted by n.
- Sample statistics are estimates of population parameters.

Not all “statistics” are statistics

In the real world, a lot of things we call “statistics” are actually parameters.

- Barry Bonds’ batting average is based on every at-bat during the season, so it is actually a parameter.

Not all “statistics” are statistics

In the real world, a lot of things we call “statistics” are actually parameters.

- Barry Bonds’ batting average is based on every at-bat during the season, so it is actually a parameter.
- Eli Manning’s pass completion percentage is also a parameter, because it is based on all of his passes in games.

Not all “statistics” are statistics

In the real world, a lot of things we call “statistics” are actually parameters.

- Barry Bonds’ batting average is based on every at-bat during the season, so it is actually a parameter.
- Eli Manning’s pass completion percentage is also a parameter, because it is based on all of his passes in games.
- The number of deaths on a highway in a year is a parameter, not a statistic.

Making good estimates

Again, for our sample statistics to be good estimates of population parameters, our sample must be:

- **representative**: the makeup of the sample should reflect the makeup of the population of interest.

Making good estimates

Again, for our sample statistics to be good estimates of population parameters, our sample must be:

- **representative**: the makeup of the sample should reflect the makeup of the population of interest.
- **random**: the chance of any two members of the population being in the sample must be equal.

Bad estimation strategies

Our statistics won’t be any good if we have a bad sample.

- If we care about all voters in the United States, we can’t use a sample of voters from Memphis.

Bad estimation strategies

Our statistics won't be any good if we have a bad sample.

- If we care about all voters in the United States, we can't use a sample of voters from Memphis.
- If we want to find out about all Ole Miss students, we can't take a sample from people in the student section at an Ole Miss basketball game.

Bad estimation strategies

Our statistics won't be any good if we have a bad sample.

- If we care about all voters in the United States, we can't use a sample of voters from Memphis.
- If we want to find out about all Ole Miss students, we can't take a sample from people in the student section at an Ole Miss basketball game.
- If we want to find out about pop singers from 15 to 30, we can't just talk to the ones who have dated Carson Daly.

Uses of statistics

We use statistics for two major reasons:

- Estimating population parameters. (For example, estimating the percentage of all voters who are Democrats.)

Uses of statistics

We use statistics for two major reasons:

- Estimating population parameters. (For example, estimating the percentage of all voters who are Democrats.)
- Testing hypotheses about populations. (For example, finding out whether Democratic voters were more likely to vote for Al Gore than Republicans.)

Two types of statistics

- **Descriptive statistics** are like population parameters; they simply describe the sample.

Two types of statistics

- **Descriptive statistics** are like population parameters; they simply describe the sample.
- **Inferential statistics** allow us to make generalizations (or inferences) about populations. Both uses of statistics in the previous slide are inferential uses of statistics.