

Section 2: Data & Measurement

Michael Gill
Gov 50

September 21, 2011

Outline

- (0) Questions/Concerns
- (1) Review of Lecture
- (2) Measurement
- (3) Sampling

Last Week

Last week, we talked about theories, which posit causal relationships between *concepts*. The goal of science more generally is to ___ these theories using data.

Last Week

Last week, we talked about theories, which posit causal relationships between *concepts*. The goal of science more generally is to **test** these theories using data.

Last Week

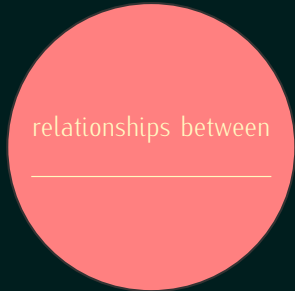
Last week, we talked about theories, which posit causal relationships between *concepts*. The goal of science more generally is to **test** these theories using data.

Theories

Last Week

Last week, we talked about theories, which posit causal relationships between *concepts*. The goal of science more generally is to **test** these theories using data.

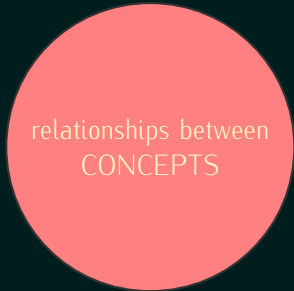
Theories



Last Week

Last week, we talked about theories, which posit causal relationships between *concepts*. The goal of science more generally is to **test** these theories using data.

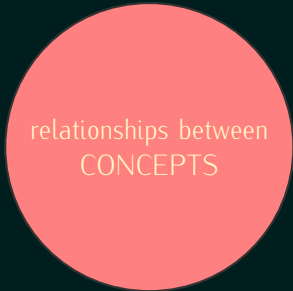
Theories



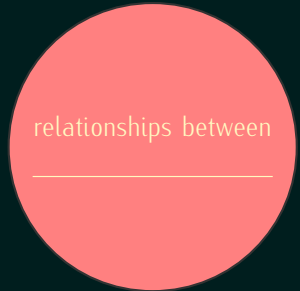
Last Week

Last week, we talked about theories, which posit causal relationships between *concepts*. The goal of science more generally is to **test** these theories using data.

Theories



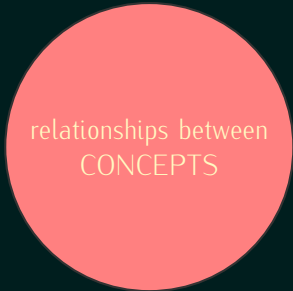
Data



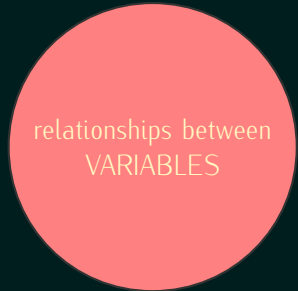
Last Week

Last week, we talked about theories, which posit causal relationships between *concepts*. The goal of science more generally is to **test** these theories using data.

Theories

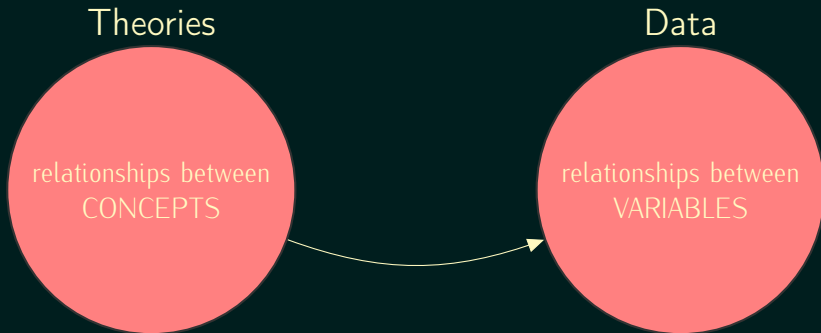


Data



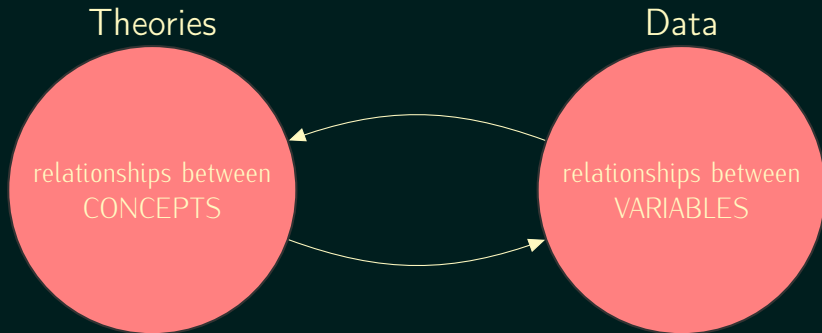
Last Week

Last week, we talked about theories, which posit causal relationships between *concepts*. The goal of science more generally is to **test** these theories using data.



Last Week

Last week, we talked about theories, which posit causal relationships between *concepts*. The goal of science more generally is to **test** these theories using data.



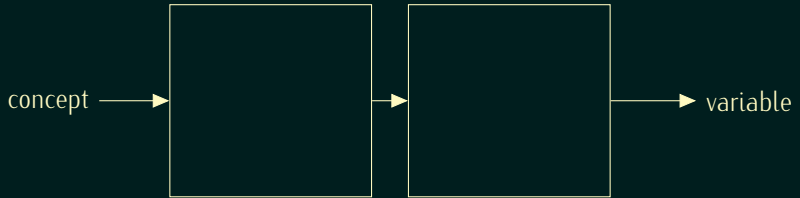
MEASUREMENT

Measurement

Measurement is the act of turning a concept into a variable. For instance, it's how we go from the concept of "being liberal" to numbers in a spreadsheet. Here is the basic process:

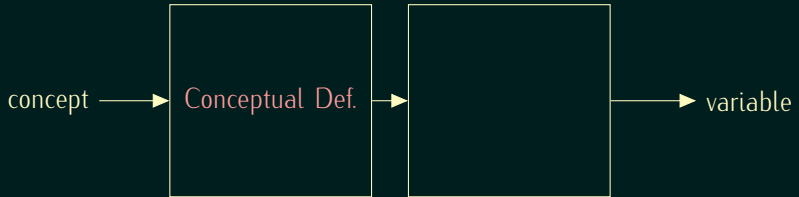
Measurement

Measurement is the act of turning a concept into a variable. For instance, it's how we go from the concept of "being liberal" to numbers in a spreadsheet. Here is the basic process:



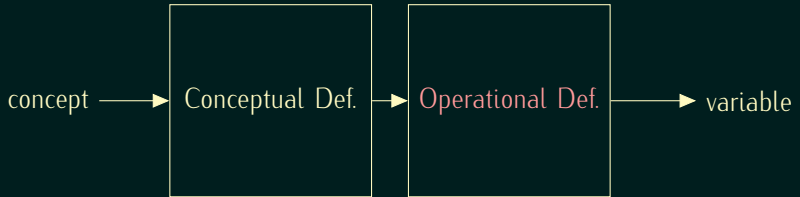
Measurement

Measurement is the act of turning a concept into a variable. For instance, it's how we go from the concept of "being liberal" to numbers in a spreadsheet. Here is the basic process:



Measurement

Measurement is the act of turning a concept into a variable. For instance, it's how we go from the concept of "being liberal" to numbers in a spreadsheet. Here is the basic process:



Measurement

Conceptual definitions help us to be explicit about what our theories are talking about. That way, we know we are measuring the right thing.

Measurement

Conceptual definitions help us to be explicit about what our theories are talking about. That way, we know we are measuring the right thing.

Conceptual definition:

Measurement

Conceptual definitions help us to be explicit about what our theories are talking about. That way, we know we are measuring the right thing.

Conceptual definition: The concept of _____ is defined as the extent to which _____ exhibit the characteristic of _____.

Some Examples:

1. What would be a good definition for the following concepts?
 - being politically informed (unit: _____)

Some Examples:

1. What would be a good definition for the following concepts?
 - being politically informed (unit: citizen)

Some Examples:

1. What would be a good definition for the following concepts?
 - being politically informed (unit: citizen)

 - being economically developed (unit: _____)

Some Examples:

1. What would be a good definition for the following concepts?
 - being politically informed (unit: citizen)
 - being economically developed (unit: country)

An Example: Political Violence

Let's say you are interested in how violence affects the political environment. For instance, you might want to know if political violence is an effective strategy for terrorist groups, or you might care if attacks increase the public's support of their leaders.

An Example: Political Violence

Let's say you are interested in how violence affects the political environment. For instance, you might want to know if political violence is an effective strategy for terrorist groups, or you might care if attacks increase the public's support of their leaders. To do this, you first have to figure **what you mean by violence**:

An Example: Political Violence

Let's say you are interested in how violence affects the political environment. For instance, you might want to know if political violence is an effective strategy for terrorist groups, or you might care if attacks increase the public's support of their leaders. To do this, you first have to figure what you mean by violence:

1. Conceptual definition of violence:

- Unit of Analysis:
- Level of the unit:
- Characteristic:

2. Operational definition:

An Example: Political Violence

Let's say you are interested in how violence affects the political environment. For instance, you might want to know if political violence is an effective strategy for terrorist groups, or you might care if attacks increase the public's support of their leaders. To do this, you first have to figure what you mean by violence:

1. Conceptual definition of violence:

- Unit of Analysis: **district, country, or region**
- Level of the unit:
- Characteristic:

2. Operational definition:

An Example: Political Violence

Let's say you are interested in how violence affects the political environment. For instance, you might want to know if political violence is an effective strategy for terrorist groups, or you might care if attacks increase the public's support of their leaders. To do this, you first have to figure what you mean by violence:

1. Conceptual definition of violence:

- Unit of Analysis: district, country, or region
- Level of the unit: individual or aggregate
- Characteristic:

2. Operational definition:

An Example: Political Violence

Let's say you are interested in how violence affects the political environment. For instance, you might want to know if political violence is an effective strategy for terrorist groups, or you might care if attacks increase the public's support of their leaders. To do this, you first have to figure what you mean by violence:

1. Conceptual definition of violence:

- Unit of Analysis: district, country, or region
- Level of the unit: individual or aggregate
- Characteristic: deaths from particular group

2. Operational definition:

An Example: Political Violence

Let's say you are interested in how violence affects the political environment. For instance, you might want to know if political violence is an effective strategy for terrorist groups, or you might care if attacks increase the public's support of their leaders. To do this, you first have to figure what you mean by violence:

1. Conceptual definition of violence:

- Unit of Analysis: district, country, or region
- Level of the unit: individual or aggregate
- Characteristic: deaths from particular group

2. Operational definition: If > 1000 deaths from particular group; civil wars; bombs dropped, etc.

Ecological Fallacy

Ecological fallacy is a type of fallacy in which a generalization about individuals is based on aggregate data about a group.

Ecological fallacy is a type of fallacy in which a generalization about individuals is based on aggregate data about a group.

Ecological fallacy is a type of fallacy in which a generalization about individuals is based on aggregate data about a group.

Ecological fallacy is a type of fallacy in which a generalization about individuals is based on aggregate data about a group.

Ecological fallacy is a type of fallacy in which a generalization about individuals is based on aggregate data about a group.

Ecological fallacy is a type of fallacy in which a generalization about individuals is based on aggregate data about a group.

Ecological fallacy is a type of fallacy in which a generalization about individuals is based on aggregate data about a group.

Ecological fallacy is a type of fallacy in which a generalization about individuals is based on aggregate data about a group.

Ecological fallacy is a type of fallacy in which a generalization about individuals is based on aggregate data about a group.

Ecological Fallacy

What is it?

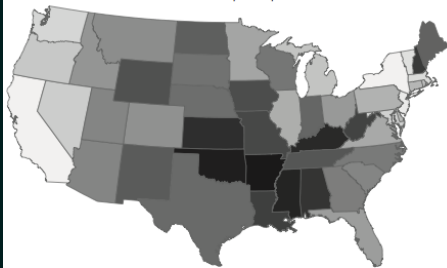
Ecological Fallacy

What is it?

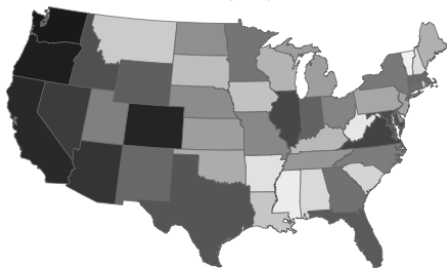
Basically, this is the assumption that associations at the **aggregate level** must also hold true at the **individual level**...

Ecological Fallacy

Wal-Marts per capita

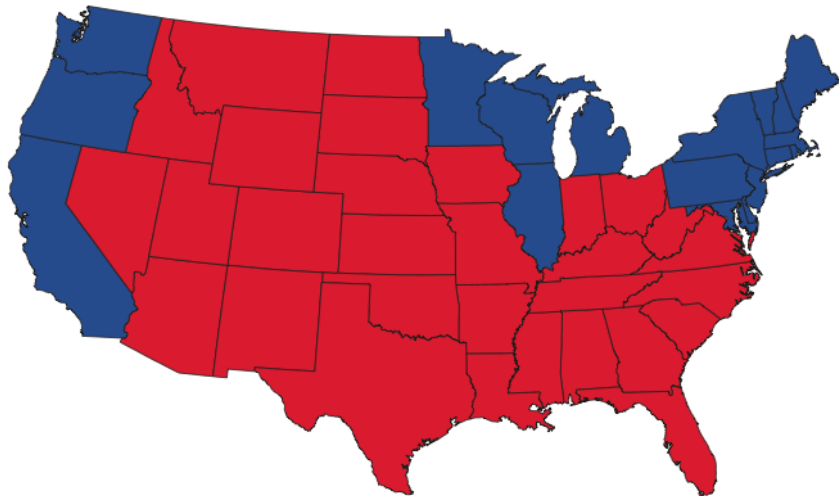


Starbucks per capita

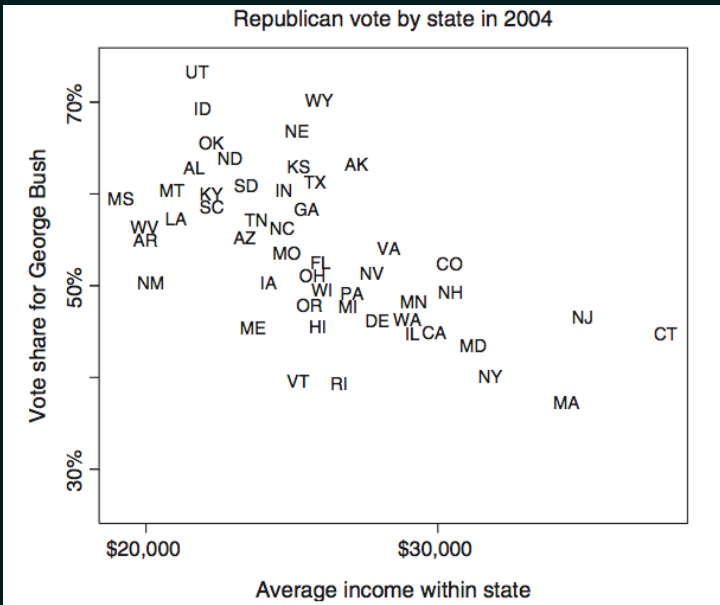


Richer States Vote Democratic...

2004 election



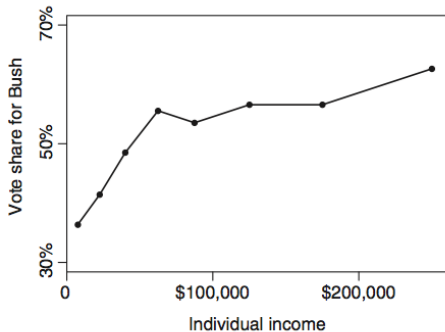
Richer States Vote Democratic...



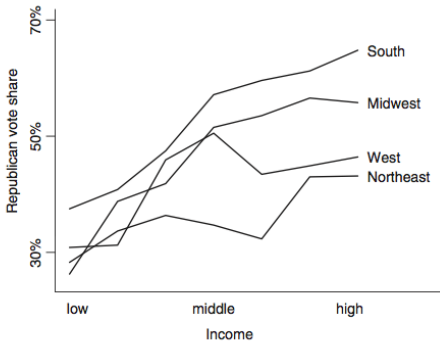
But Richer People Vote Republican...

But Richer People Vote Republican...

Bush vote in 2004 by income



2006 House exit polls



Another Example: Berkeley Gender Bias Case

Another Example: Berkeley Gender Bias Case

Evidence of (the concept of) gender bias?

Another Example: Berkeley Gender Bias Case

Evidence of (the concept of) gender bias?

	Applicants	Admitted
Men	8442	44%
Women	4321	33%

Another Example: Berkeley Gender Bias Case

Evidence of (the concept of) gender bias?

	Applicants	Admitted
Men	8442	44%
Women	4321	33%

Another Example: Berkeley Gender Bias Case

Evidence of (the concept of) gender bias?

	Applicants	Admitted
Men	8442	44%
Women	4321	33%



Another Example: Berkeley Gender Bias Case

Evidence of (the concept of) gender bias?

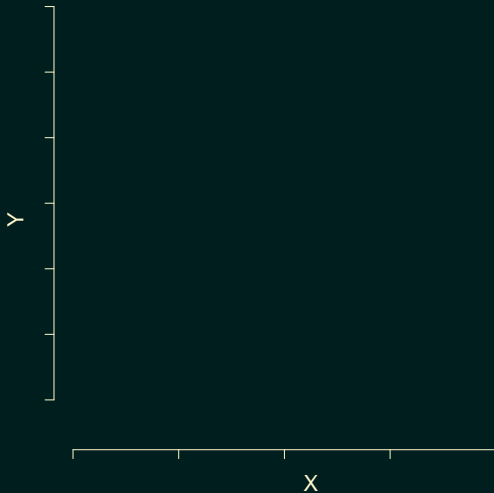
	Applicants	Admitted
Men	8442	44%
Women	4321	33%



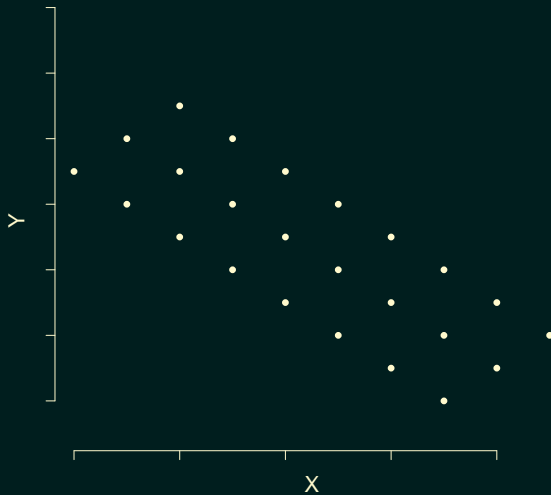
Dept.	Male App	Male Admit.	Female App.	Female Admit.
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	272	6%	341	7%

What is the relationship between X and Y ?

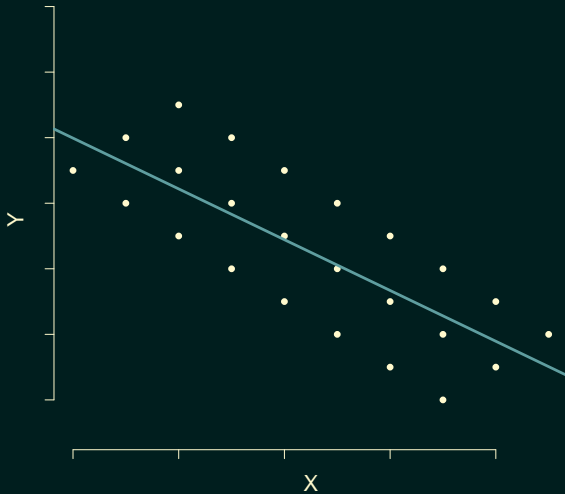
Relationship Between X and Y



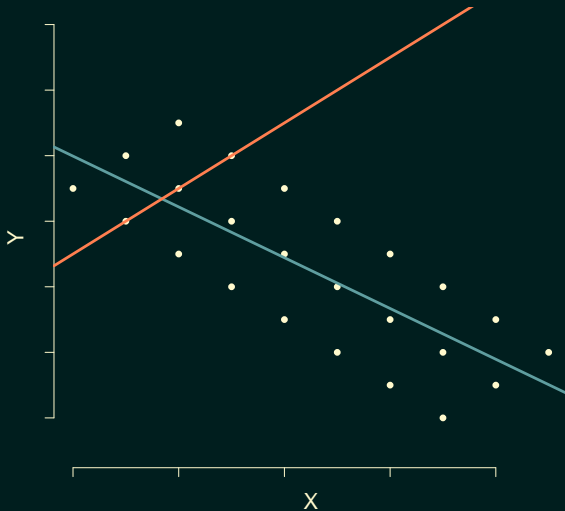
Relationship Between X and Y



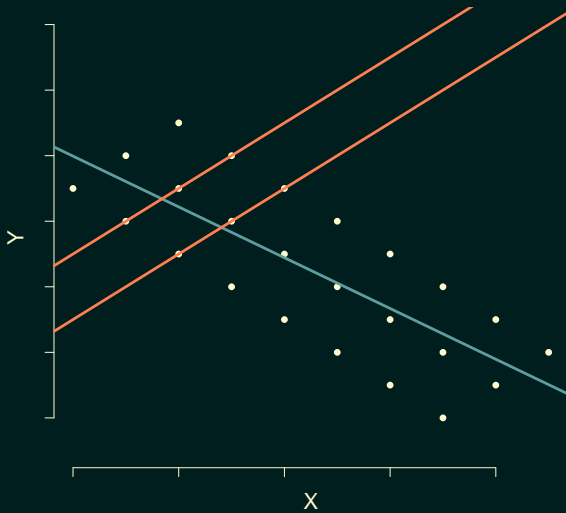
Relationship Between X and Y



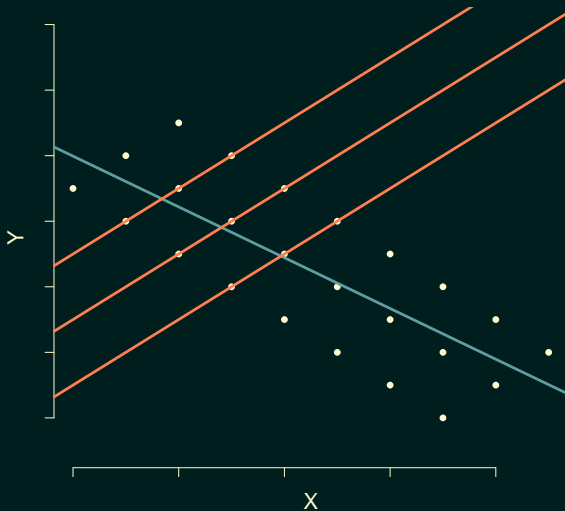
Relationship Between X and Y



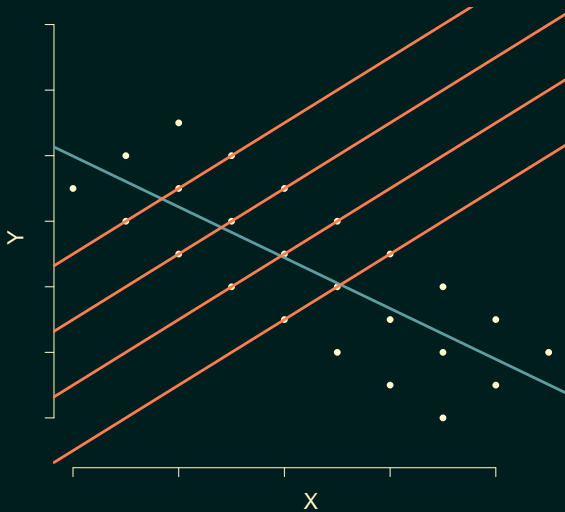
Relationship Between X and Y



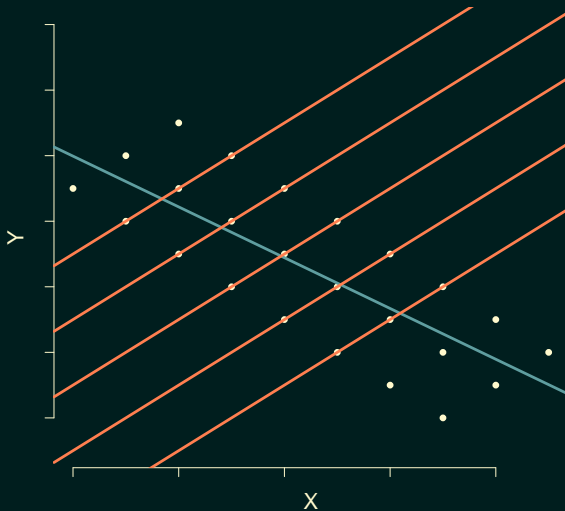
Relationship Between X and Y



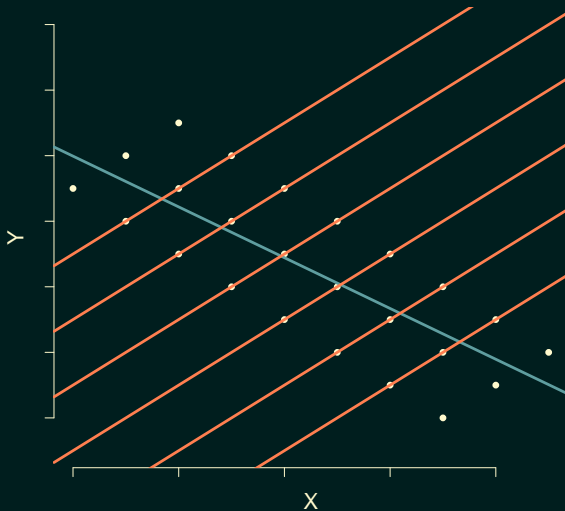
Relationship Between X and Y



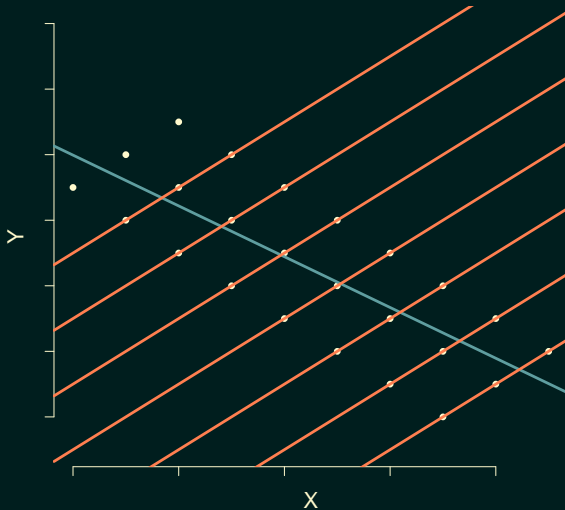
Relationship Between X and Y



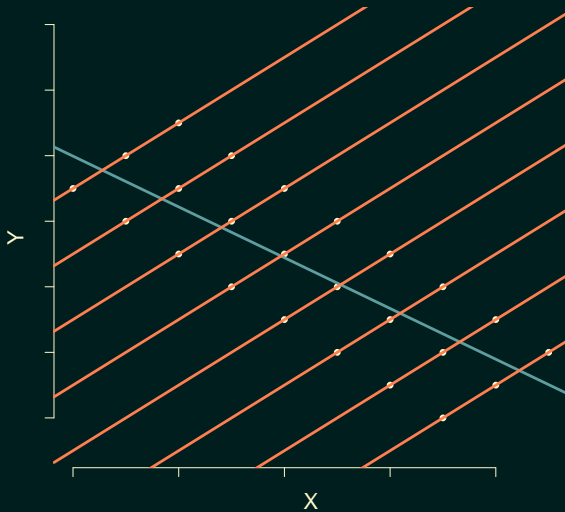
Relationship Between X and Y



Relationship Between X and Y



Relationship Between X and Y



Measurement

Operational definitions describe how we convert the concept into the variable numerical codes.

Measurement

Operational definitions describe how we convert the concept into the variable numerical codes. What could be a simple example of a **operational definition** for a **sex variable**?

Measurement

Operational definitions describe how we convert the concept into the variable numerical codes. What could be a simple example of a operational definition for a sex variable?

- 1 if a person is Female;

Measurement

Operational definitions describe how we convert the concept into the variable numerical codes. What could be a simple example of a operational definition for a sex variable?

- 1 if a person is Female;
- 0 if a person is Male.

Measurement (or, Mad Libs!)

1. ***Systematic measurement error*** occurs when the _____ definition fails to match the _____ definition in a _____ manner. Lack of systematic error is called _____.

Measurement (or, Mad Libs!)

1. ***Systematic measurement error*** occurs when the **operational** definition fails to match the _____ definition in a _____ manner. Lack of systematic error is called _____.

Measurement (or, Mad Libs!)

1. ***Systematic measurement error*** occurs when the **operational** definition fails to match the **conceptual** definition in a _____ manner. Lack of systematic error is called _____.

Measurement (or, Mad Libs!)

1. ***Systematic measurement error*** occurs when the **operational** definition fails to match the **conceptual** definition in a **systematic** manner. Lack of systematic error is called _____.

Measurement (or, Mad Libs!)

1. ***Systematic measurement error*** occurs when the **operational** definition fails to match the **conceptual** definition in a **systematic** manner. Lack of systematic error is called **validity**.

Measurement (or, Mad Libs!)

1. ***Systematic measurement error*** occurs when the **operational** definition fails to match the **conceptual** definition in a **systematic** manner. Lack of systematic error is called **validity**.
2. ***Random measurement error*** occurs when _____ factors affect the measurement. Lack of this random error is called _____.

Measurement (or, Mad Libs!)

1. ***Systematic measurement error*** occurs when the **operational** definition fails to match the **conceptual** definition in a **systematic** manner. Lack of systematic error is called **validity**.
2. ***Random measurement error*** occurs when **temporary** or **haphazard** factors affect the measurement. Lack of this random error is called _____.

Measurement (or, Mad Libs!)

1. ***Systematic measurement error*** occurs when the **operational** definition fails to match the **conceptual** definition in a **systematic** manner. Lack of systematic error is called **validity**.
2. ***Random measurement error*** occurs when **temporary** or **haphazard** factors affect the measurement. Lack of this random error is called **reliability**.

SAMPLING

What is sampling?



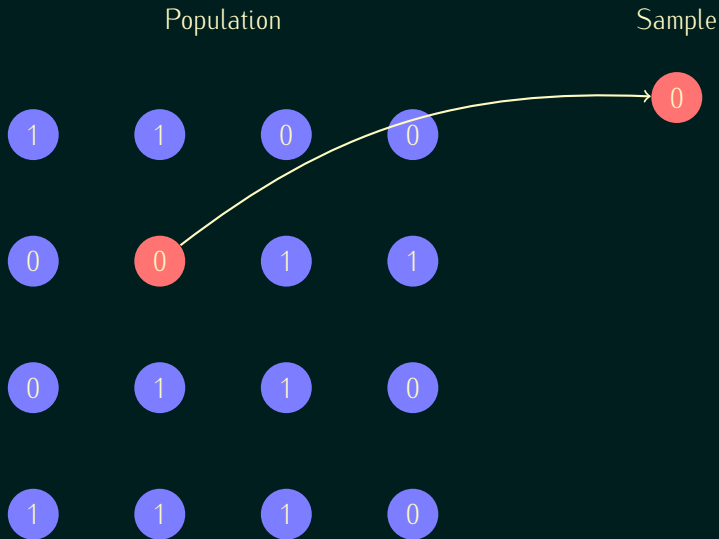
Random Sampling Error

Population

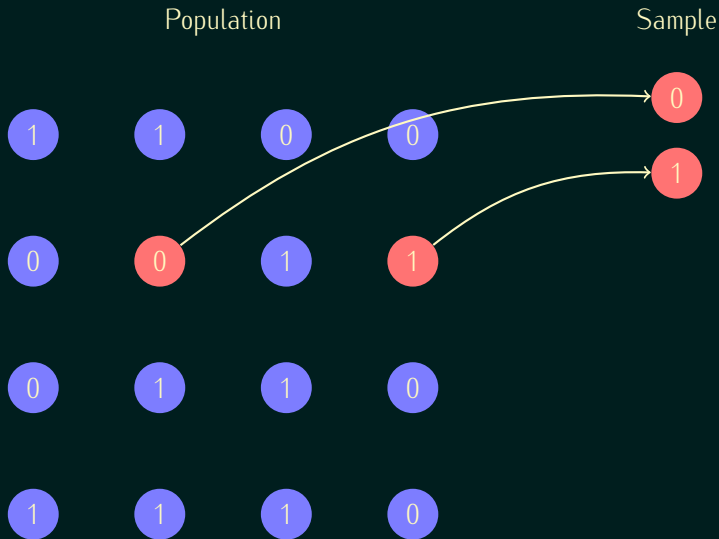
Sample

1	1	0	0
0	0	1	1
0	1	1	0
1	1	1	0

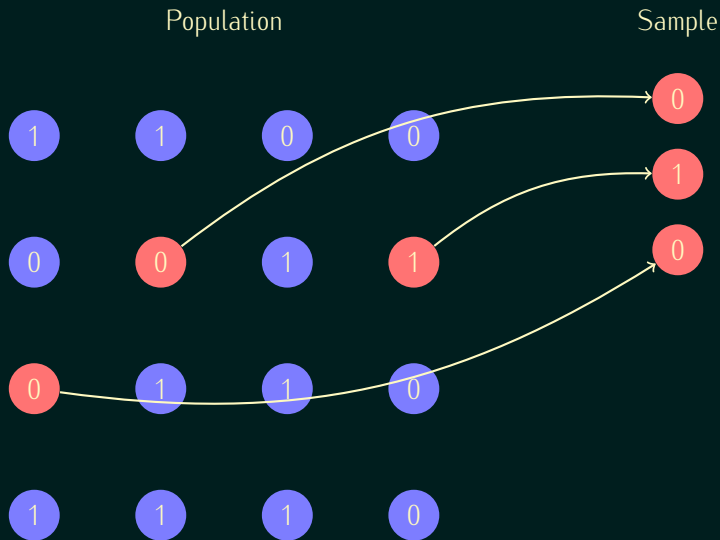
Random Sampling Error



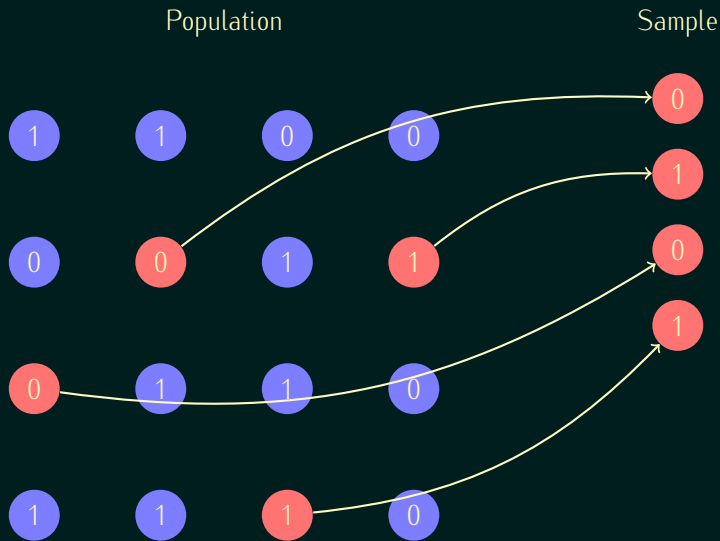
Random Sampling Error



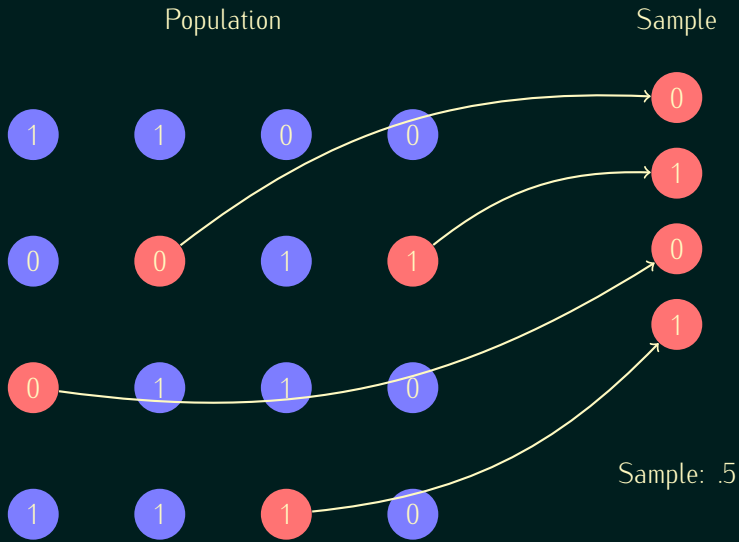
Random Sampling Error



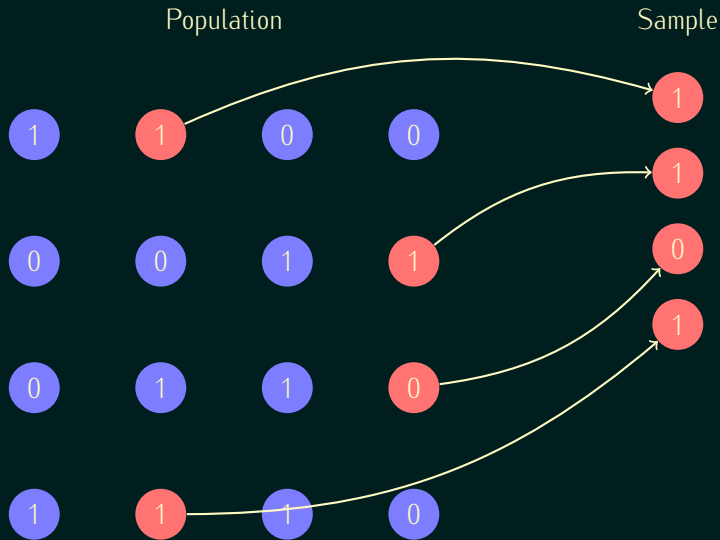
Random Sampling Error



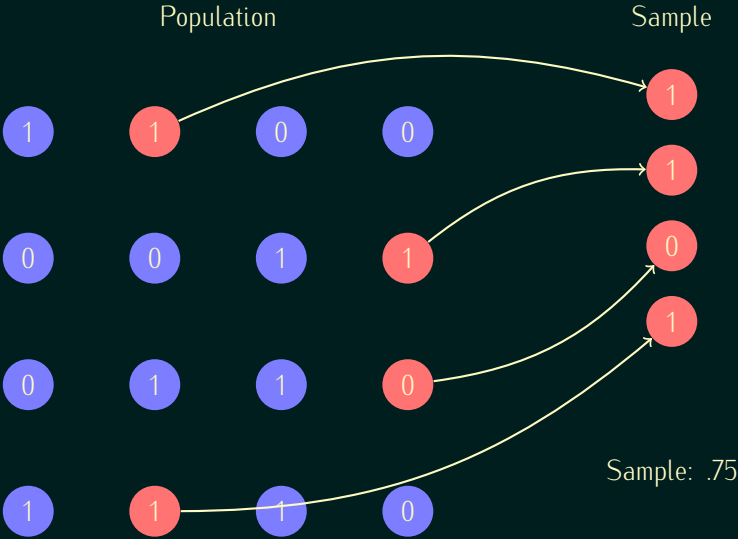
Random Sampling Error



Random Sampling Error



Random Sampling Error



Random Sampling Error

$$\text{Random Sampling Error} = \frac{\text{component}}{\text{component}}$$

Random Sampling Error

$$\text{Random Sampling Error} = \frac{\text{variation}}{\text{component}}$$

Random Sampling Error

$$\text{Random Sampling Error} = \frac{\text{variation}}{\text{sample size}} \frac{\text{component}}{\text{component}}$$

Random Sampling Error

1. Suppose an enterprising Gov50 student filled up two bathtubs with 1,000 marbles each. In bathtub 1, she used 1 red marble and 999 blue marbles. In bathtub 2, she used 500 of each color. If she took a sample of 50 marbles from both bathtubs, which would have less RSE? Why?

500 red, 500 blue

1 red, 999 blue

Random Sampling Error

1. Suppose an enterprising Gov50 student filled up two bathtubs with 1,000 marbles each. In bathtub 1, she used 1 red marble and 999 blue marbles. In bathtub 2, she used 500 of each color. If she took a sample of 50 marbles from both bathtubs, which would have less RSE? Why?

500 red, 500 blue

1 red, 999 blue