Social Data Science

SOCIOL 114 Winter 2025

Lecture 1: Asking Research Questions

Learning goals for this course

By the end of this course, you will be able to

- visualize economic inequality with graphs that summarize survey data
- connect theories about inequality to quantitative empirical evidence
- evaluate the effects of hypothetical interventions to reduce inequality
- conduct data analysis using the R programming language

What makes a good quantitative research question?

1. a unit of analysis

► a row of your dataset

- 1. a unit of analysis
 - ► a row of your dataset
- 2. an outcome
 - ► a variable with a value for each unit

1. a unit of analysis

- a row of your dataset
- 2. an outcome

a variable with a value for each unit

- 3. a target population
 - ▶ a set of units about whom to infer
 - clear who is included and who is not

1. a unit of analysis

- a row of your dataset
- 2. an outcome
 - a variable with a value for each unit
- 3. a target population
 - ▶ a set of units about whom to infer
 - clear who is included and who is not
- 4. potential for surprising results

A good project may have a very simple question

What proportion of children born in large U.S. cities in 1998–2000 was ever evicted from their home from birth to age 15?

What proportion of children born in large U.S. cities in 1998–2000 was ever evicted from their home from birth to age 15?

unit of analysis

target population



What proportion of children born in large U.S. cities in 1998–2000 was ever evicted from their home from birth to age 15?

unit of analysis

a child

target population

► outcome

What proportion of children born in large U.S. cities in 1998–2000 was ever evicted from their home from birth to age 15?

unit of analysis

a child

target population

children born in large U.S. cities in 1998–2000

outcome

What proportion of children born in large U.S. cities in 1998–2000 was ever evicted from their home from birth to age 15?

unit of analysis

a child

target population

children born in large U.S. cities in 1998–2000

(and subgroups by race and income)

outcome

What proportion of children born in large U.S. cities in 1998–2000 was ever evicted from their home from birth to age 15?

unit of analysis

a child

target population

children born in large U.S. cities in 1998–2000

(and subgroups by race and income)

outcome

evicted from home between birth and age 15

Lundberg & Donnelly 2019



Lundberg & Donnelly 2019



H19. We are also interested in some of the problems that families face making ends meet. In the past 12 months, did you do any of the following because there wasn't enough money?

Lundberg & Donnelly 2019



H19. We are also interested in some of the problems that families face making ends meet. In the past 12 months, did you do any of the following because there wasn't enough money?

	[YES	NO
H19E.	(In the past 12 months), were you evicted from your home or		
	apartment for not paying the rent or mortgage?	1	2

Lundberg & Donnelly 2019



H19. We are also interested in some of the problems that families face making ends meet. In the past 12 months, did you do any of the following because there wasn't enough money?

		YES	NO
H19E.	(In the past 12 months), were you evicted from your home or		
	apartment for not paying the rent or mortgage?	1	2

we filled in missing values with regression

Lundberg & Donnelly 2019



H19. We are also interested in some of the problems that families face making ends meet. In the past 12 months, did you do any of the following because there wasn't enough money?

		YES	NO
H19E.	(In the past 12 months), were you evicted from your home or		
	apartment for not paying the rent or mortgage?	1	2

- we filled in missing values with regression
- we gathered responses across years



1. a unit of analysis

- a row of your dataset
- 2. an outcome
 - a variable with a value for each unit
- 3. a target population
 - ▶ a set of units about whom to infer
 - clear who is included and who is not
- 4. potential for surprising results

Course logistics

soc114.github.io

What about this course makes you feel anxious?

What about this course makes you feel excited?

Appendix: Causal research question

Describe a population

What is the proportion employed among U.S. resident women ages 21–35?

Describe a population

What is the proportion employed among U.S. resident women ages 21–35?

Woman 1 Woman 2 Woman 3 Woman 4 Describe a population

What is the proportion employed among U.S. resident women ages 21–35?

Employed?

Woman 1	1
Woman 2	0
Woman 3	1
Woman 4	1

What is the proportion employed among U.S. resident women ages 21–35, comparing mothers to non-mothers? What is the proportion employed among U.S. resident women ages 21–35, comparing mothers to non-mothers?

	Employed?		Employed?
Mother 1	0	Non-Mother 1	1
Mother 2	0	Non-Mother 2	0
Mother 3	0	Non-Mother 3	1
Mother 4	1	Non-Mother 4	1

Woman 1 Woman 2 Woman 3 Woman 4

	Would be
	employed if
	a mother?
	Y(1)
Woman 1	0
Woman 2	0
Woman 3	0
Woman 4	1

	Would be	Would be
	employed if	employed if
	a mother?	a non-mother?
	Y(1)	Y(0)
Woman 1	0	1
Woman 2	0	0
Woman 3	0	1
Woman 4	1	1

	Would be	Would be	
	employed if	employed if	Causal
	a mother?	a non-mother?	effect
	Y(1)	Y(0)	$\frac{Y(1)-Y(0)}{}$
Woman 1	0	1	-1
Woman 2	0	0	0
Woman 3	0	1	-1
Woman 4	1	1	0

Describe population subgroups

What is the proportion employed among U.S. resident women ages 21–35, comparing mothers to non-mothers?

	Employed?		Employed?
Mother 1	0	Non-Mother 1	1
Mother 2	0	Non-Mother 2	0
Mother 3	0	Non-Mother 3	1
Mother 4	1	Non-Mother 4	1



Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? Y(1)	Would be employed if a non-mother? Y(0)	Causal effect Y(1) - Y(0)
Woman 1	0	1	-1
Woman 2	0	0	0
Woman 3	0	1	-1
Woman 4	1	1	0

Descriptive

among

across

difference

for those who

Descriptive	Causal
among	causes
across	affects
difference	produces
for those who	impacts

Descriptive	Murky Middle	Causal
among	associated with	causes
across	leads to	affects
difference	predicts	produces
for those who		impacts

Descriptive	Murky Middle	Causal
among	associated with	causes
across	leads to	affects
difference	predicts	produces
for those who		impacts
		^

verbs

↑ not verbs		↑ verbs
for those who		impacts
difference	predicts	produces
across	leads to	affects
among	associated with	causes
Descriptive	Murky Middle	Causal

Descriptive	Murky Middle	Causal
among	associated with	causes
across	leads to	affects
difference	predicts	produces
for those who		impacts
↑ not verbs		↑ verbs
C		(

Statements "predictor verb outcome" are often causal

(analysis needs a DAG!)

Descriptive	Murky Middle	Causal
among	associated with	causes
across	leads to	affects
difference	predicts	produces
for those who		impacts
↑ not verbs		↑ verbs

Statements "predictor verb outcome"(analysis needsare often causala DAG!)

Statements "among subpopulation, mean outcome" are often descriptive

What proportion of children born in large U.S. cities in 1998–2000 who lived in public housing at age 9 would have been evicted at age 9–15 if they had lived in a private rental?

Example: Effect of public housing on eviction Lundberg et al. 2021

What proportion of children born in large U.S. cities in 1998–2000 who lived in public housing at age 9 would have been evicted at age 9–15 if they had lived in a private rental?

unit of analysis

target population



Lundberg et al. 2021

What proportion of children born in large U.S. cities in 1998–2000 who lived in public housing at age 9 would have been evicted at age 9–15 if they had lived in a private rental?

unit of analysis

a child

target population



Lundberg et al. 2021

What proportion of children born in large U.S. cities in 1998–2000 who lived in public housing at age 9 would have been evicted at age 9–15 if they had lived in a private rental?

unit of analysis

► a child

- target population
 - children born in large U.S. cities in 1998–2000 who lived in public housing at age 9
- outcome

Lundberg et al. 2021

What proportion of children born in large U.S. cities in 1998–2000 who lived in public housing at age 9 would have been evicted at age 9–15 if they had lived in a private rental?

unit of analysis

► a child

- target population
 - children born in large U.S. cities in 1998–2000 who lived in public housing at age 9
- outcome
 - evicted from home between age 9 and 15

We took the same dataset:



We took the same dataset:



For every kid in public housing, we estimated the rate of eviction among kids who **looked like them** but who live in a private rental We took the same dataset:



For every kid in public housing, we estimated the rate of eviction among kids who **looked like them** but who live in a private rental

We **assumed** those rates would have happened to our target population in the absence of public housing

Effect of public housing on eviction

