

Studying Social Inequality with Data Science

INFO 3370 / 5371
Spring 2024

Course Summary

Goals of the course

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

What we did in this course

1. Population Sampling
2. Working with Data
3. Describing Inequality
4. Reducing Inequality

leading to your own final projects



116

dy elstern zu dir vñ vat
mit in all geman
re oach hin zu einem fram
vñ nach mit der rist dar an
Das ref der vil rann matt
vñ arz er an den fram gelugt
wazzen mer dat genug
in em pach vil groz

Inz dem heiten fram flos
das vecel vñ haint vñ all d scha
Genug zetzenbchen heten gr. d.
Dw zogt auf s mit groz dy
vñ amalech dy hadenjast
vñ witten mit in freiten
Is dy amalechiten
degund näh en gem den me.

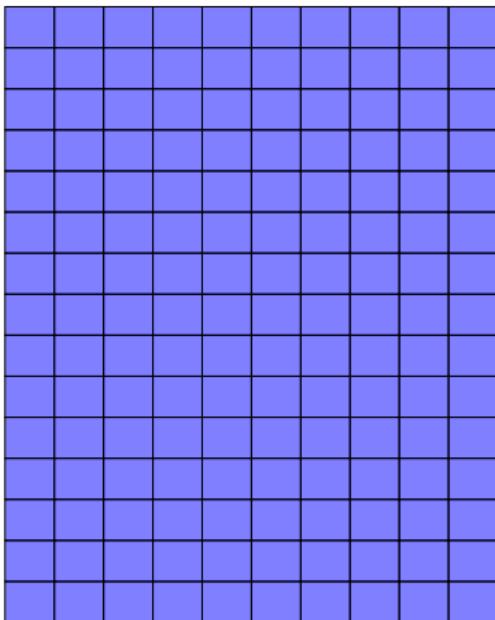
The Israelites
Collecting Manna
from Heaven
Austria
About 1400–1410

Unknown artist
Getty Open Content Program
getty.edu/art/collection/object/105T6R

Population Sampling

Full Count Enumeration

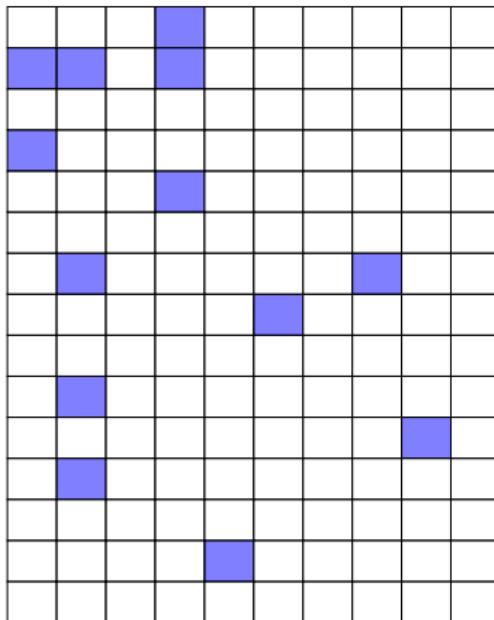
Back of Room



Front of Room

Probability Sample

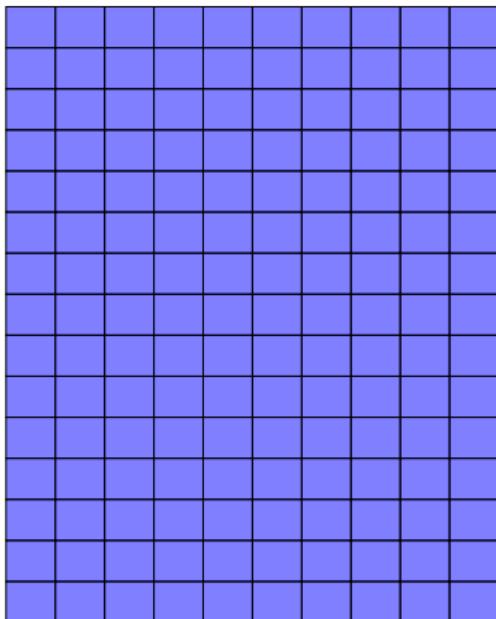
Back of Room



Front of Room

Full Count Enumeration

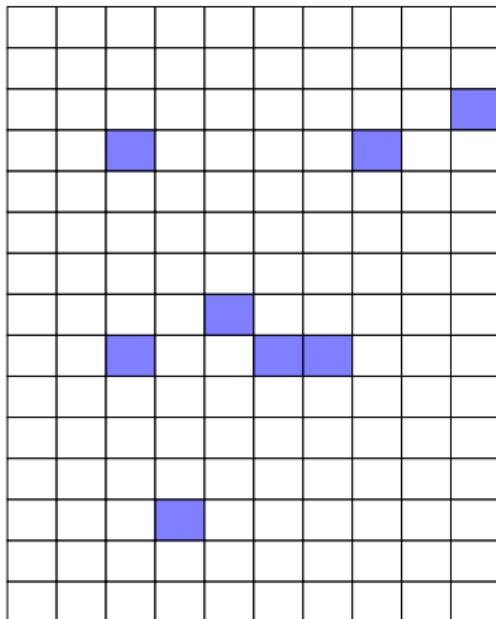
Back of Room



Front of Room

Probability Sample

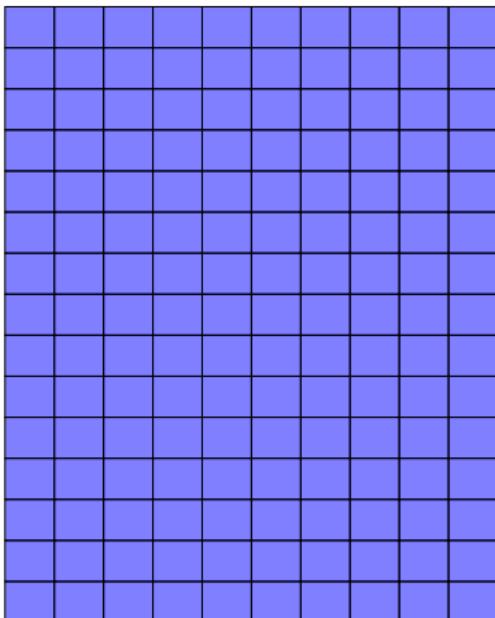
Back of Room



Front of Room

Full Count Enumeration

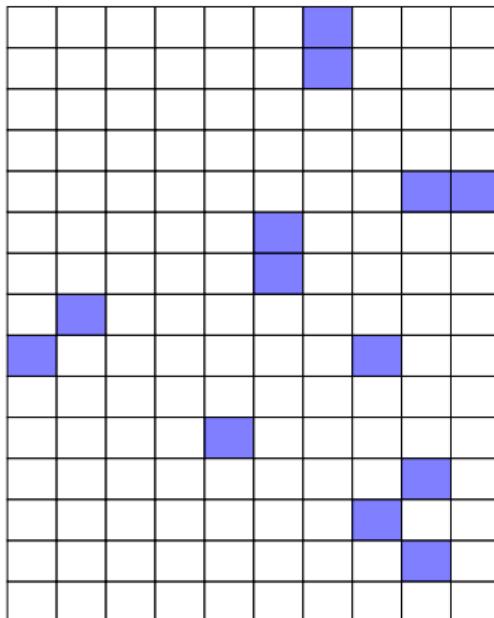
Back of Room



Front of Room

Probability Sample

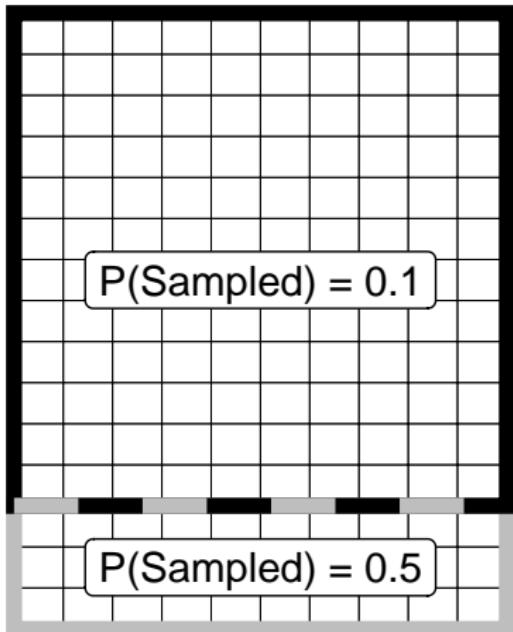
Back of Room



Front of Room

Sample Design

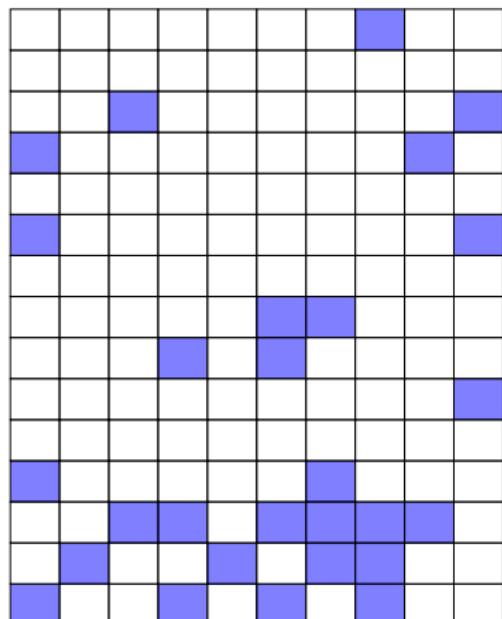
Back of Room



Front of Room

Sample

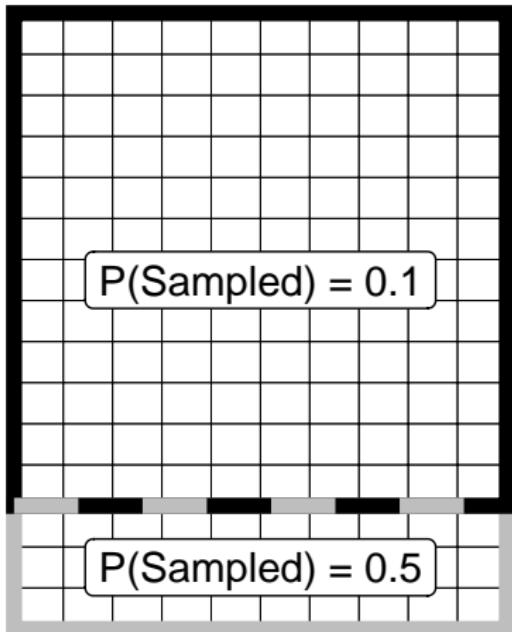
Back of Room



Front of Room

Sample Design

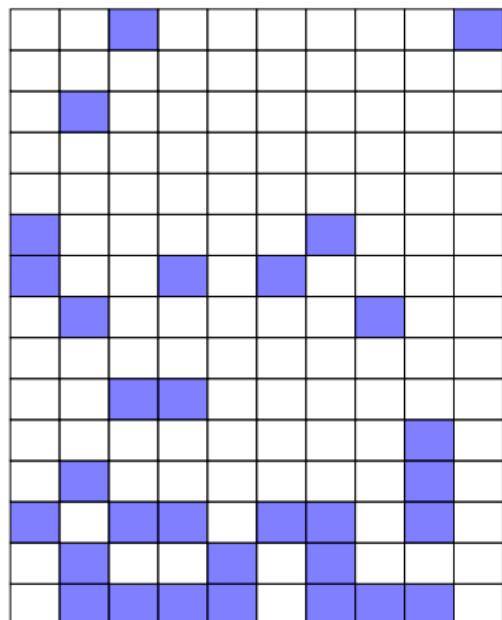
Back of Room



Front of Room

Sample

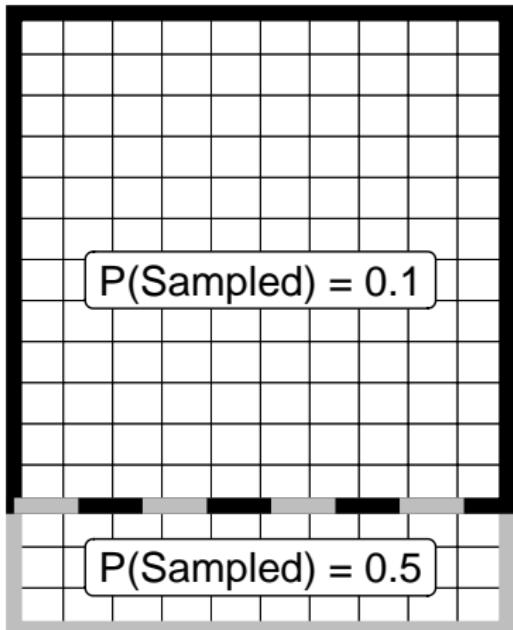
Back of Room



Front of Room

Sample Design

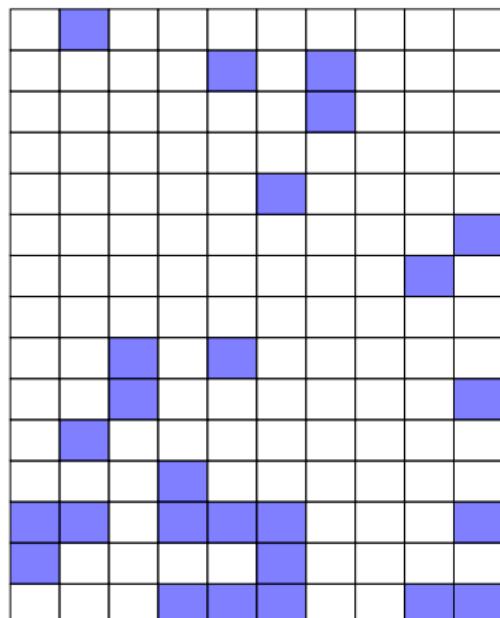
Back of Room



Front of Room

Sample

Back of Room



Front of Room

Working with Data

Working with Data: Finding Data

LOG IN | REGISTER | IPUMS.ORG  IPUMS CPS

CURRENT POPULATION SURVEY

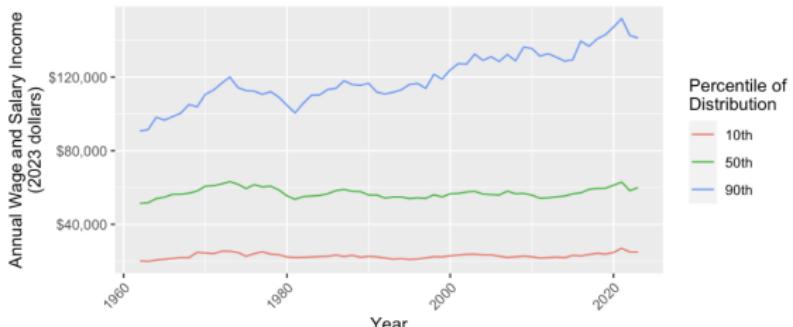
HOME | SELECT DATA | MY DATA | SUPPORT

Working with Data: Transforming Data

Topics > Working with Data > Data transformation

Data transformation

This exercise examines how income inequality has changed over time in the U.S. We will measure inequality by the 10th, 50th, and 90th percentiles of wage and salary income from 1962 to 2022. We expect that this exercise may take more time than one discussion session.¹ You will begin by downloading data and end by making this graph.



On this page

[Data access](#)

[Prepare R environment](#)

[Get familiar with our dataset](#)

[filter\(\) to cases of interest](#)

[group_by\(\) and summarize\(\) for subpopulation summaries](#)

[pivot_longer\(\) to reshape data](#)

[left_join\(\) an inflation adjustment](#)

[mutate\(\) to adjust for inflation](#)

[ggplot\(\) to visualize](#)

[All together](#)

[Finished early?](#)

Working with Data: Transforming Data

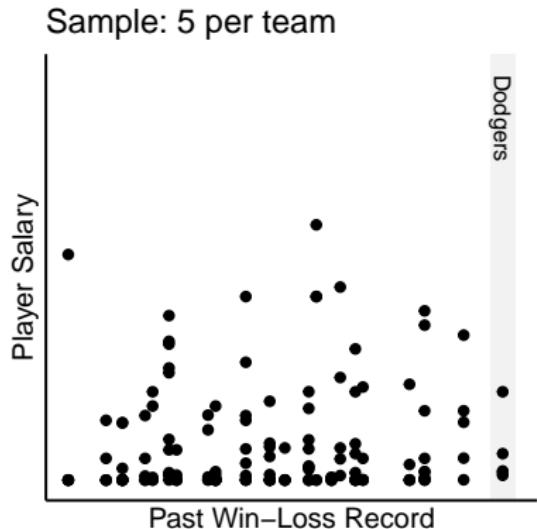
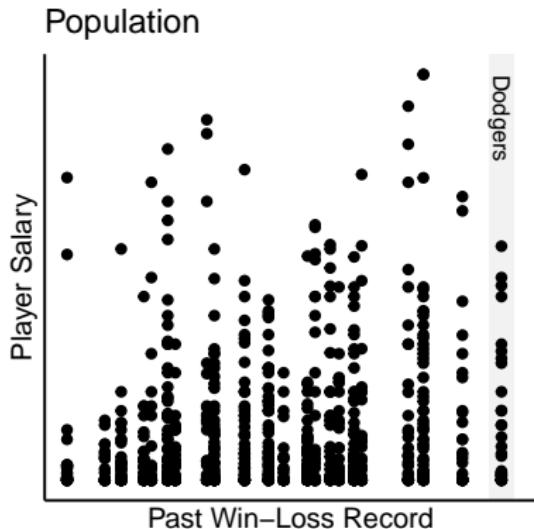
Questions of the form:

Among ____, what is the mean of ____?

(population inference from a sample)

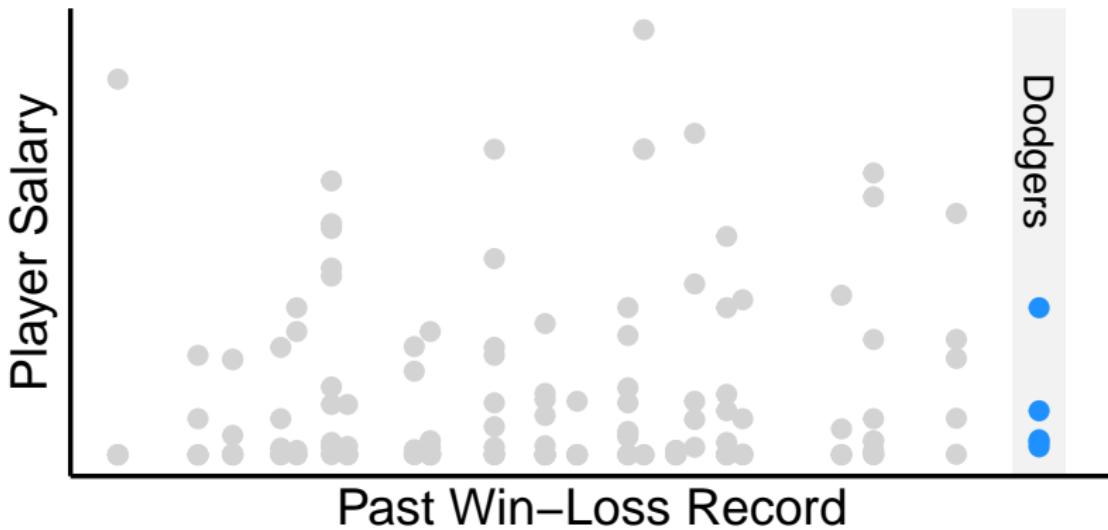
Working with Data: Statistical Learning

With only the sample, estimate the mean salary of the Dodgers



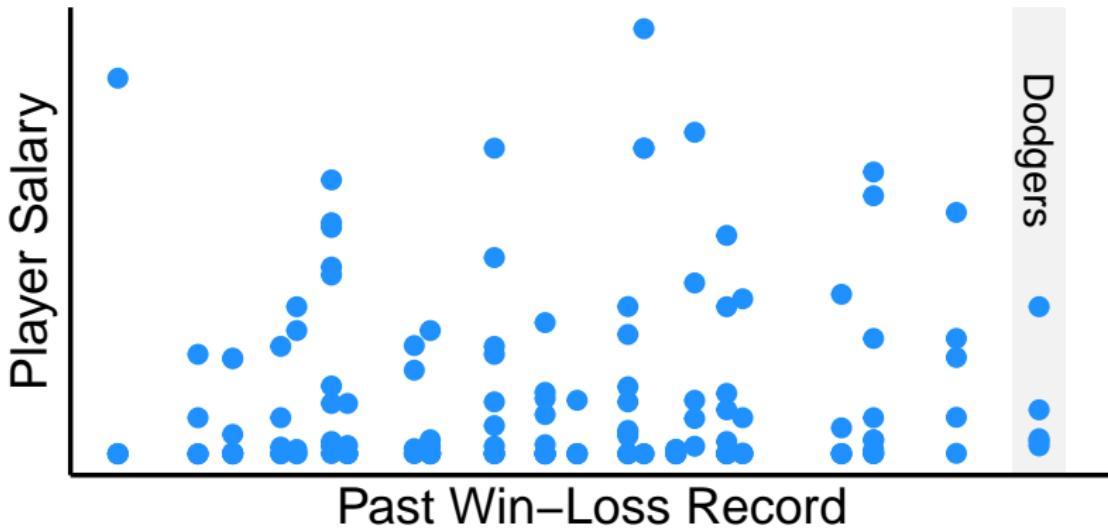
Working with Data: Statistical Learning

Estimator 1: Subgroup sample mean



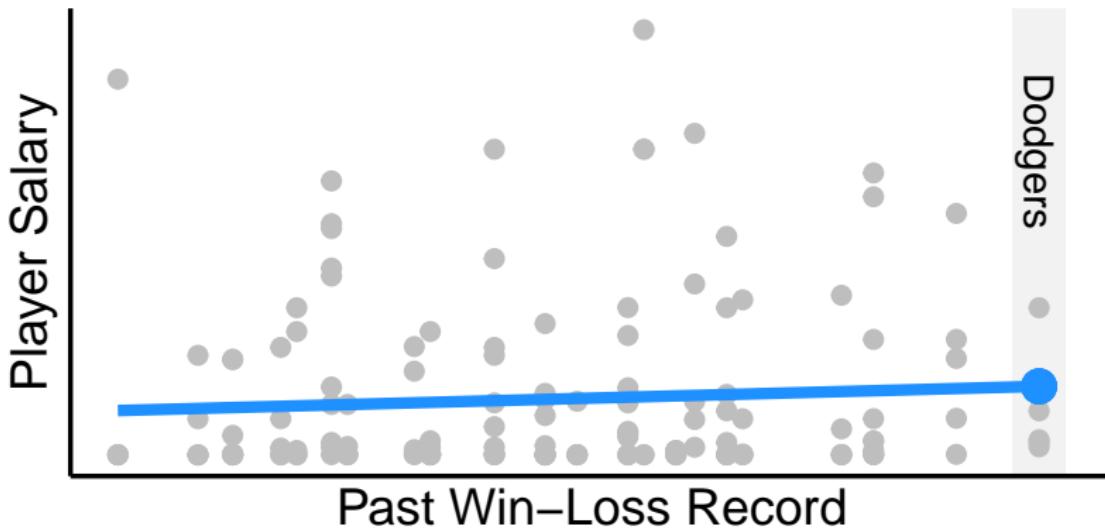
Working with Data: Statistical Learning

Estimator 2: Full sample mean

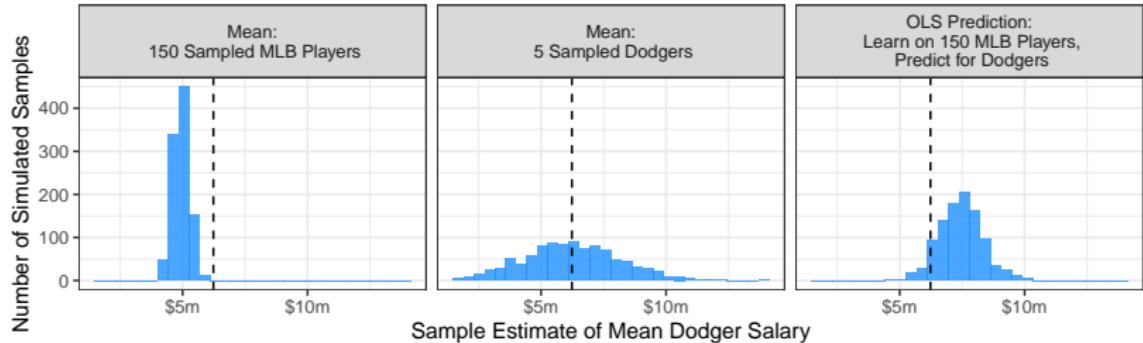
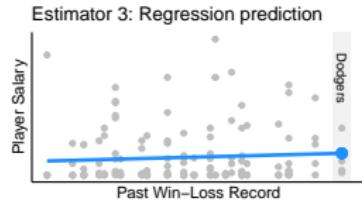
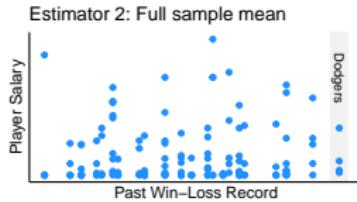
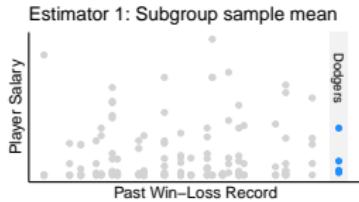


Working with Data: Statistical Learning

Estimator 3: Regression prediction



Working with Data: Statistical Learning



Working with Data: Statistical Learning

Learning Set

Holdout Set

Case 1
Case 2
Case 3
Case 4
Case 5

	Respondent Education	Parent Education	Grandparent Education	Sex	Race	Grandparent Income	Parent Income
Case 1							
Case 2							
Case 3							
Case 4							
Case 5							

Learn a prediction function →

Respondent Income

Predict for new cases →

?	?	?

Describing Inequality

Describing inequality: We practiced descriptive research questions

- ▶ race: the median white household has \$5.48 for each \$1 held by the median Black household
 - ▶ and we discussed political origins

Describing inequality: We practiced descriptive research questions

- ▶ race: the median white household has \$5.48 for each \$1 held by the median Black household
 - ▶ and we discussed political origins
- ▶ gender: “the ratio of women’s pay to men’s pay increased from 0.61 to 0.83 between 1970 and 2018”
 - ▶ from England et al. 2020, in Problem Set 2

Describing inequality: We practiced descriptive research questions

- ▶ race: the median white household has \$5.48 for each \$1 held by the median Black household
 - ▶ and we discussed political origins
- ▶ gender: “the ratio of women’s pay to men’s pay increased from 0.61 to 0.83 between 1970 and 2018”
 - ▶ from England et al. 2020, in Problem Set 2
- ▶ class: professional workers have higher incomes if their parents were also professional workers
 - ▶ Laurison & Friedman 2024, in Problem Set 4

Reducing Inequality

Reducing Inequality: Moral Arguments

What would make a distribution just?

Reducing Inequality: Moral Arguments

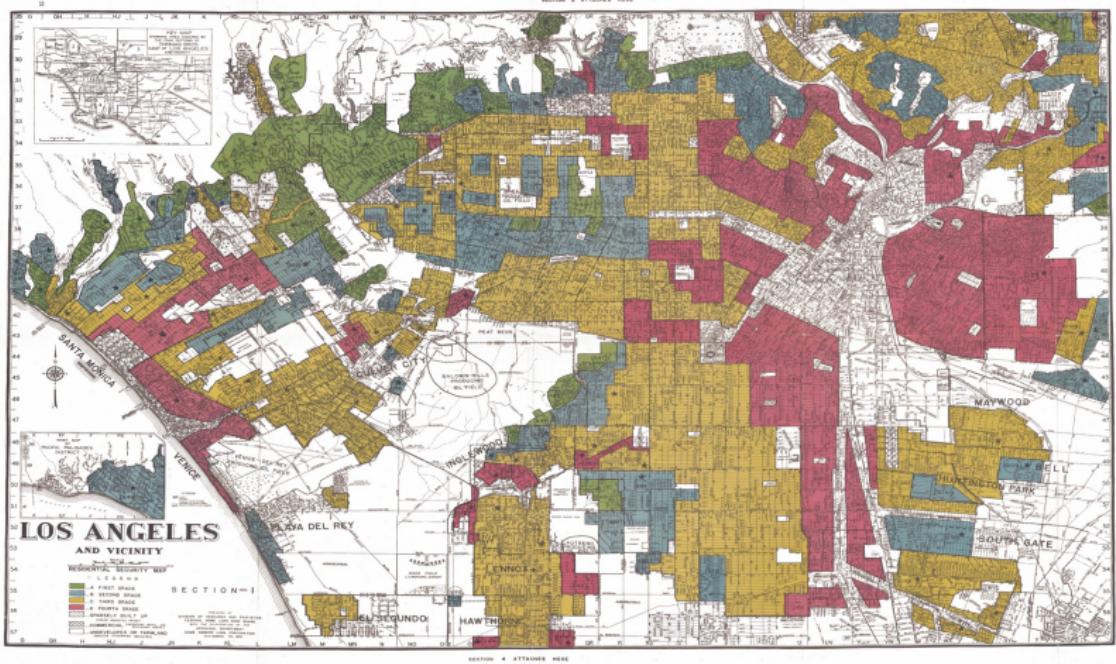
What would make a distribution just?

- We'd choose it in the original position (Rawls)

Reducing Inequality: Moral Arguments

What would make a distribution just?

- ▶ We'd choose it in the original position (Rawls)
- ▶ It arises through just acquisitions and just transfers (Nozick)



Source: Nelson and Ayers, [Mapping Inequality](#)



Source: Wikimedia



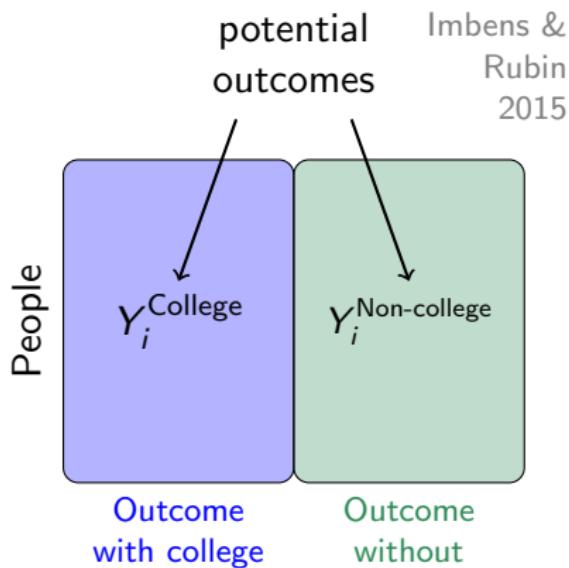
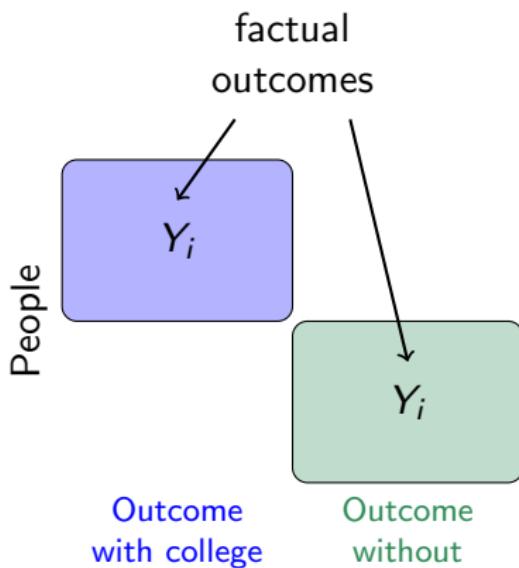
Source: County of Los Angeles

Reducing Inequality: Causal Inference

Two claims

People with
college degrees
earn more

A college degree
causes
higher earnings



Final Project

What we did in this course

1. Population Sampling
2. Working with Data
3. Describing Inequality
4. Reducing Inequality

leading to your own final projects

Feedback

1. What will you walk away now able to do?
2. What worked especially well?
3. How could this course be even better?

Course evaluations

Course evaluations

- ▶ Specific examples are especially helpful
 - ▶ Comments on course content
 - ▶ Comments on course delivery

Course evaluations

- ▶ Specific examples are especially helpful
 - ▶ Comments on course content
 - ▶ Comments on course delivery
- ▶ Specific positive experience with a TA? Tell us by name!

Thank you for a great semester!