

SOC-5811 Week 3: Quantitative research methods

Nick Graetz

University of Minnesota, Department of Sociology

9/17/2025

1/25



- ▶ **Empiricism:** social theories imply empirical claims that are falsifiable.

QUANTITATIVE METHODS

- ▶ **Empiricism:** social theories imply empirical claims that are falsifiable.
- ▶ Data doesn't speak for itself; it must be carefully interpreted, summarize, and analyzed.



QUANTITATIVE METHODS

- ▶ Statistics is about providing a concise summary of empirical data.



QUANTITATIVE METHODS

- ▶ Statistics is about providing a concise summary of empirical data.
- ▶ This involves analyzing **samples** drawn from **populations**.



SAMPLES AND POPULATIONS

Say I have a bag filled with 50 marbles of different colors. I reach in the bag and pull out 10 marbles. I see that 6 marbles are blue and 4 are red.

- What is the population and sample here?



SAMPLES AND POPULATIONS

Say I have a bag filled with 50 marbles of different colors. I reach in the bag and pull out 10 marbles. I see that 6 marbles are blue and 4 are red.

- ▶ What is the population and sample here?
- ▶ What can I say about the sample? What can I say about the population?



GENERALIZATION

Summarizing in-sample

- ▶ **Measurement:** generalizing from **observed measurements** to the **underlying constructs of interest**.



GENERALIZATION

Summarizing in-sample

- ▶ **Measurement:** generalizing from **observed measurements** to the **underlying constructs of interest**.
 - ▶ Is what I'm measuring what I actually care about?



GENERALIZATION

Summarizing in-sample

- ▶ **Measurement:** generalizing from **observed measurements** to the **underlying constructs of interest**.
 - ▶ Is what I'm measuring what I actually care about?
 - ▶ *I ask 30 people whether they graduated from college; what am I measuring?*



GENERALIZATION

Summarizing in-sample

- ▶ **Measurement:** generalizing from **observed measurements** to the **underlying constructs of interest**.
 - ▶ Is what I'm measuring what I actually care about?
 - ▶ *I ask 30 people whether they graduated from college; what am I measuring?*
 - ▶ Measurement theory: very central to psychology, education, etc.



GENERALIZATION

Summarizing in-sample

- ▶ **Measurement:** generalizing from **observed measurements** to the **underlying constructs of interest**.
 - ▶ Is what I'm measuring what I actually care about?
 - ▶ *I ask 30 people whether they graduated from college; what am I measuring?*
 - ▶ Measurement theory: very central to psychology, education, etc.
- ▶ **Generalizing from sample to population.**

GENERALIZATION

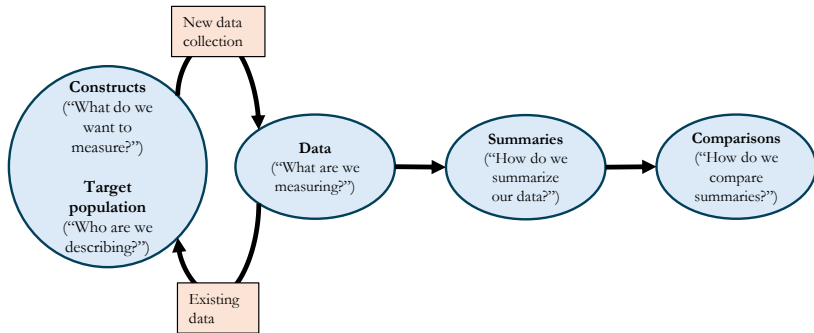
Summarizing in-sample

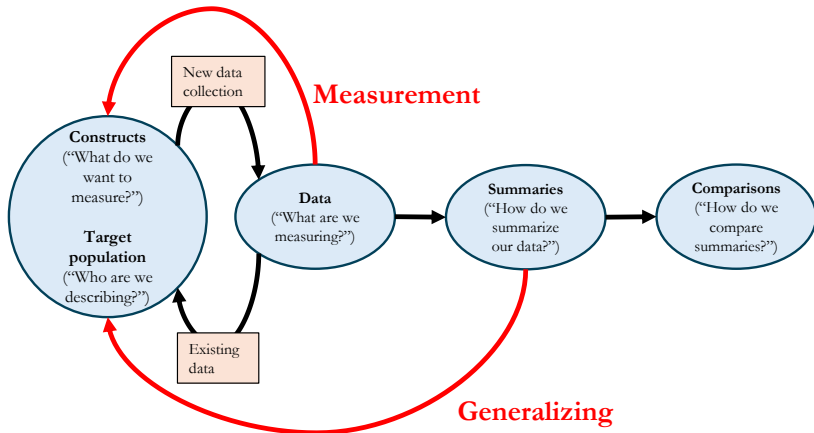
- ▶ **Measurement:** generalizing from **observed measurements** to the **underlying constructs of interest**.
 - ▶ Is what I'm measuring what I actually care about?
 - ▶ *I ask 30 people whether they graduated from college; what am I measuring?*
 - ▶ Measurement theory: very central to psychology, education, etc.
- ▶ **Generalizing from sample to population.**
 - ▶ Is what I see in the sample also true in the population?

GENERALIZATION

Summarizing in-sample

- ▶ **Measurement:** generalizing from **observed measurements** to the **underlying constructs of interest**.
 - ▶ Is what I'm measuring what I actually care about?
 - ▶ *I ask 30 people whether they graduated from college; what am I measuring?*
 - ▶ Measurement theory: very central to psychology, education, etc.
- ▶ **Generalizing from sample to population.**
 - ▶ Is what I see in the sample also true in the population?
 - ▶ *I ask 30 people whether they graduated from college; what can I say about the population from which I sampled these people?*





Generalization

- ▶ **Forecasting:** generalizing from the present to the future.



Generalization

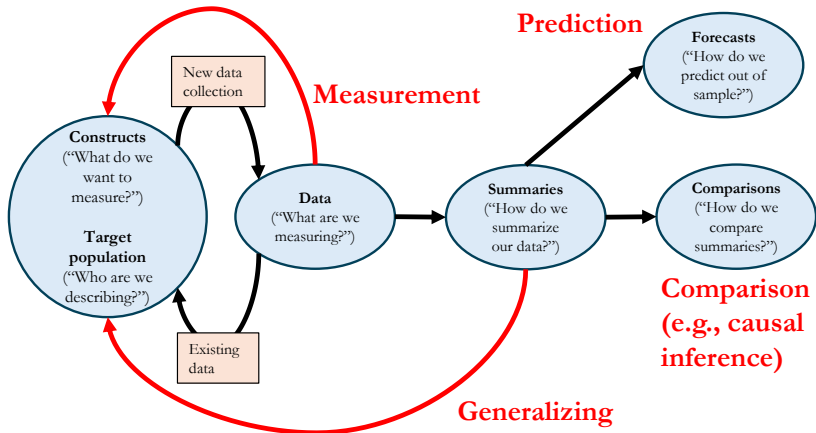
- ▶ **Forecasting:** generalizing from the present to the future.
 - ▶ *I have data on whether or not it has rained every day for the past ten years. Will it rain tomorrow?*

Generalization

- ▶ **Forecasting:** generalizing from the present to the future.
 - ▶ *I have data on whether or not it has rained every day for the past ten years. Will it rain tomorrow?*
- ▶ **Causal inference:** generalizing from treatment group to control group.

Generalization

- ▶ **Forecasting:** generalizing from the present to the future.
 - ▶ *I have data on whether or not it has rained every day for the past ten years. Will it rain tomorrow?*
- ▶ **Causal inference:** generalizing from treatment group to control group.
 - ▶ *I measure the rate of a specific disease among a group that was vaccinated for that disease and one that was not. Was the vaccine effective at reducing the disease?*



GENERALIZATION: SUMMARY

1. Measurement.
2. Generalizing from sample to population.
3. Comparison (e.g., causal inference).
4. Forecasting.



GENERALIZATION: SUMMARY

1. Measurement.
2. **Generalizing from sample to population.**
3. **Comparison (e.g., causal inference.)**
4. Forecasting.



QUANTITATIVE DATA

Data on US populations by state/region

```
## # A tibble: 6 x 7
##   msa                population percapita state region isusa rank
##   <chr>              <dbl>      <dbl> <fct> <fct> <fct> <dbl>
## 1 Washington-Arlington-Alexandria~ 5949.      47.4 11_DC 1_Nor~ Is i~    1
## 2 San Jose-Santa Clara-Sunnyvale,~ 1919.      40.4 6_CA  4_West Is i~    2
## 3 Atlanta, Georgia MSA             5545.      25.3 13_GA 3_Sou~ Is i~   15
## 4 Austin-San Marcos, Texas MSA      2271.      24.5 48_TX 3_Sou~ Is i~   21
## 5 Little Rock-North Little Rock, ~    724.      20.3 5_AR  3_Sou~ Is i~  112
## 6 Bellingham, Washington MSA        167.      20.0 53_WA 4_West Is i~  121
```


QUANTITATIVE DATA

Continuous and discrete variables

```
## # A tibble: 5 x 2
##   population region
##   <dbl> <fct>
## 1    5949. 1_Northeast
## 2    1919. 4_West
## 3    5545. 3_South
## 4    2271. 3_South
## 5     724. 3_South
```

LEVELS OF MEASUREMENT

- ▶ **Continuous variables:** infinite number of values in a given interval (e.g., time)



LEVELS OF MEASUREMENT

- ▶ **Continuous variables:** infinite number of values in a given interval (e.g., time)
- ▶ **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)



LEVELS OF MEASUREMENT

- ▶ **Continuous variables:** infinite number of values in a given interval (e.g., time)
- ▶ **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)
 - ▶ Ordinal: ordered, discrete categories (e.g., ranks in a race)



LEVELS OF MEASUREMENT

- ▶ **Continuous variables:** infinite number of values in a given interval (e.g., time)
- ▶ **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)
 - ▶ Ordinal: ordered, discrete categories (e.g., ranks in a race)
 - ▶ Nominal: unordered, discrete categories (e.g., field of study)



LEVELS OF MEASUREMENT

- ▶ **Continuous variables:** infinite number of values in a given interval (e.g., time)
- ▶ **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)
 - ▶ Ordinal: ordered, discrete categories (e.g., ranks in a race)
 - ▶ Nominal: unordered, discrete categories (e.g., field of study)
 - ▶ Binary/dichotomous: special case of nominal variable that only has two values (e.g., coin flip)

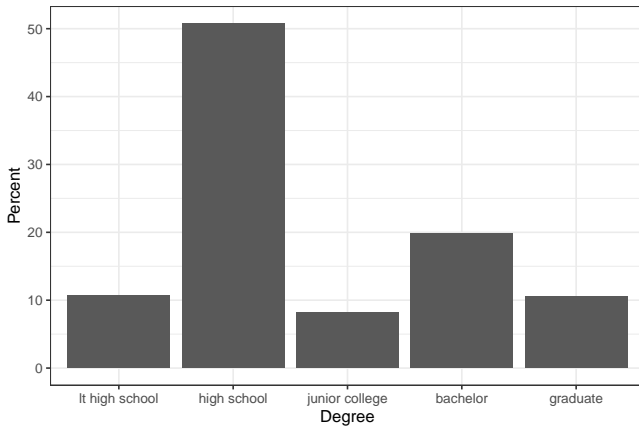
LEVELS OF MEASUREMENT

- ▶ **Continuous variables:** infinite number of values in a given interval (e.g., time)
- ▶ **Discrete variables:** limited to a specific set of outcomes (e.g., die roll)
 - ▶ Ordinal: ordered, discrete categories (e.g., ranks in a race)
 - ▶ Nominal: unordered, discrete categories (e.g., field of study)
 - ▶ Binary/dichotomous: special case of nominal variable that only has two values (e.g., coin flip)
- ▶ Measurement decisions are not always clear cut (e.g., age)

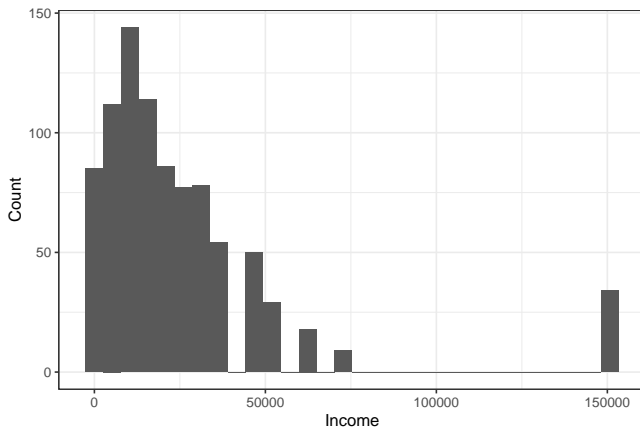
SUMMARIZING DATA

```
## # A tibble: 6 x 8
##   degree      agekdbrn  realrinc  race  age      prestg10  educ      happy
##   <fct>      <dbl+lbl> <dbl+lbl> <fct> <dbl+lbl> <dbl+lbl> <dbl+lbl> <fct>
## 1 bachelor    35      45400  white  42        61        16    very h~
## 2 bachelor    32      54480  white  63        59        16    very h~
## 3 lt high school 17        908    white  62        35         8    not to~
## 4 high school  30      45400  white  55        39        12    pretty~
## 5 graduate    30      54480  white  59        72        19    pretty~
## 6 high school  20      8512.  other  34        35        14    pretty~
```


SUMMARIZING DATA



SUMMARIZING DATA

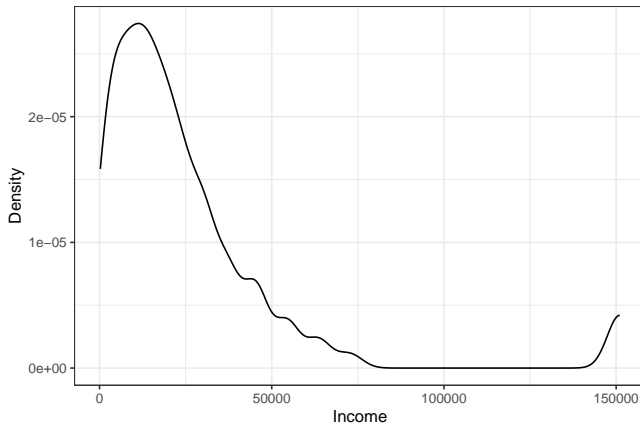


SUMMARIZING DATA

- ▶ Histograms of continuous data: why do we have to bin?
- ▶ How do we choose bins?

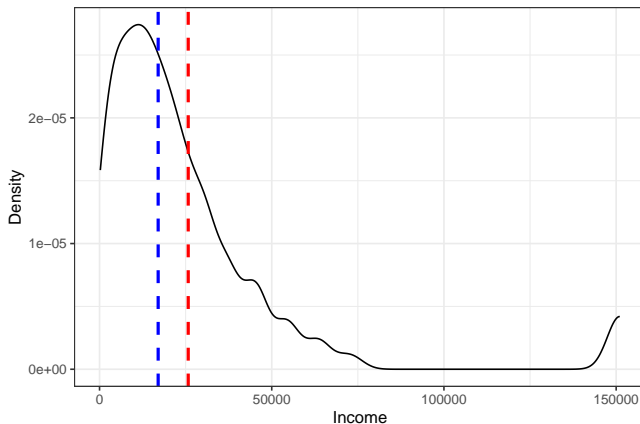


SUMMARIZING DATA



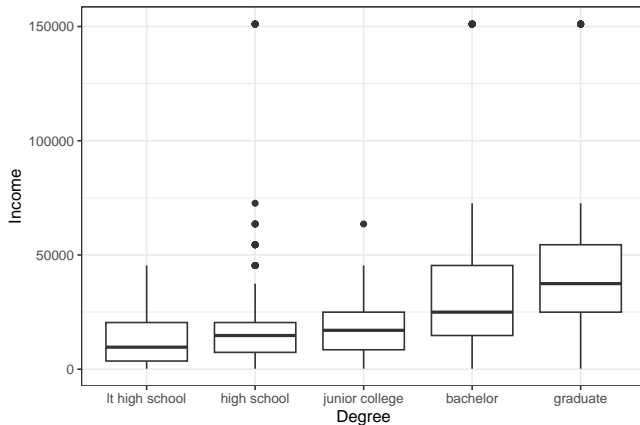
SUMMARIZING DATA

Add measures of central tendency: mean and median



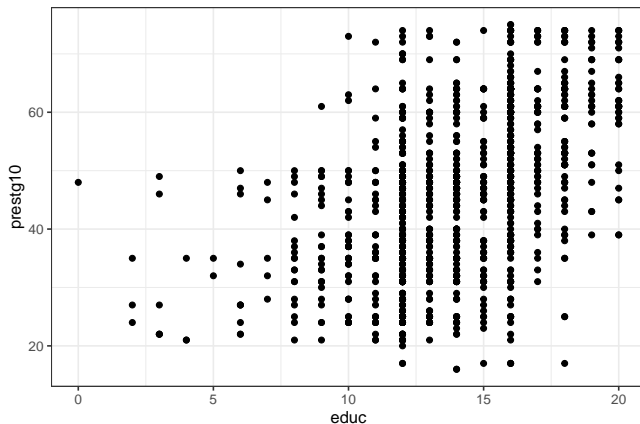
UNIVARIATE SUMMARY STATISTICS

Examine distribution:



BIVARIATE RELATIONSHIPS

Examine bivariate relationship:



BIVARIATE RELATIONSHIPS

Correlation matrix:

##	educ	prestg10	realrinc	agekdbrn
## educ	1.00	0.49	0.35	0.42
## prestg10	0.49	1.00	0.30	0.37
## realrinc	0.35	0.30	1.00	0.22
## agekdbrn	0.42	0.37	0.22	1.00

BIVARIATE RELATIONSHIPS

Cross-tabulations table:

##		race		
##	degree	white	black	other
##	lt high school	0.06	0.02	0.03
##	high school	0.36	0.09	0.05
##	junior college	0.06	0.01	0.00
##	bachelor	0.15	0.03	0.02
##	graduate	0.08	0.01	0.01

STATS AND SOCIAL SCIENCE THEORY

How do we create models?

- ▶ Deterministic processes
- ▶ Stochastic processes

