Political Science 210: Introduction to Empirical Methods

Week 3: Data and Measurement

Types of Variables

- Categorical
  - Nominal
    - Different "types" with no ordering (e.g. type of electoral system)
  - Ordinal
    - Different "types" with a particular order: (e.g. level of education)
    - Unclear distance between intervals
- Numeric
  - Discrete
    - Data you can *count*: (e.g. number of political parties in a system)
    - Interval between each level is equal
  - Continuous
    - Data you can *measure*: (e.g. percentage of a population)
    - (Possibly) infinite number of values between intervals

Types of Variables

	Variable type	Does the order matter?	Equal distance between intervals?	Infinite number of possible values?
Categorical	Nominal	N	N	N
	Ordinal	Y	N	N
Numeric	Discrete	Y	Y	N
	Continuous	Y	Y	Y

- The type of variable we choose reflects how we think about our concept.
- In the concepts below, is there any sort of order in the variables we might use? Do we know the distance between the intervals? Does adding complexity to the variable improve our understanding of the concept or weaken it?
  - Partisanship?
  - Political ideology?
  - Strength of democracy?

Types of Variables

Let's say you have two theories about the relationship between higher education and political ideology. What type of variable would you choose for the dependent and independent variables in each of the theories?

- As people learn more with each level of education, they become more aware of other perspectives and grow more culturally liberal.

- When people go to college, they socialize with students who tend to hold liberal views and are likely to become liberals.

Error and Bias

Broadly, there are two possible sources of error when measuring our variables:

- Reliability: Does your measurement strategy obtain the same result every time you measure?
- Validity: Does your measurement strategy measure what you intend to measure? Does it measure the concept ("construct validity") or what it claims to measure ("measurement validity")?

There are two types of error produced by measurement issues:

- When you introduce a large amount of random error, you have an *unreliable measure*.
- If you are regularly taking the wrong measurement in a consistent direction, you have *systemic bias*.

Error and Bias

A. Theranos claims its technology can run blood tests on just a drop of blood, making it far less invasive. Tests show measured blood sugar levels get the correct value on average, though each individual sample might be off by as much as 15%.

B. Using only four simple questions, researchers can find out whether a respondent's personality fits the authoritarian profile. These questions categorize people correctly nearly every time, and repeated applications of the quiz usually reveals consistent results.

C. Researchers exposed subjects to political ads while connecting them to electrodes that measure pulse and perspiration. In repeated tests, people's baseline values were the same. However, researchers later learned that the equipment was poorly calibrated, so that the pulse levels were all exactly 5% too high.

D. A researcher wants to know if wealthier people vote at higher rates. To answer this question she wants to use the money a person spends on eating out each month to measure their wealth, as reported by survey respondents. A look at this data shows that this amount has high variability. Additionally, a second researcher points out that the spending figures might be systematically too low, as people tend to underestimate their spending.

Place each letter in the corresponding grid in the table:

	Low validity	High validity
Low reliability		
High reliability		

Error and Bias

A. Theranos claims its technology can run blood tests on just a drop of blood, making it far less invasive. Tests show measured blood sugar levels get the correct value on average, though each individual sample might be off by as much as 15%.

B. Using only four simple questions, researchers can find out whether a respondent's personality fits the authoritarian profile. These questions categorize people correctly nearly every time, and repeated applications of the quiz usually reveals consistent results.

C. Researchers exposed subjects to political ads while connecting them to electrodes that measure pulse and perspiration. In repeated tests, people's baseline values were the same. However, researchers later learned that the equipment was poorly calibrated, so that the pulse levels were all exactly 5% too high.

D. A researcher wants to know if wealthier people vote at higher rates. To answer this question she wants to use the money a person spends on eating out each month to measure their wealth, as reported by survey respondents. A look at this data shows that this amount has high variability. Additionally, a second researcher points out that the spending figures might be systematically too low, as people tend to underestimate their spending.

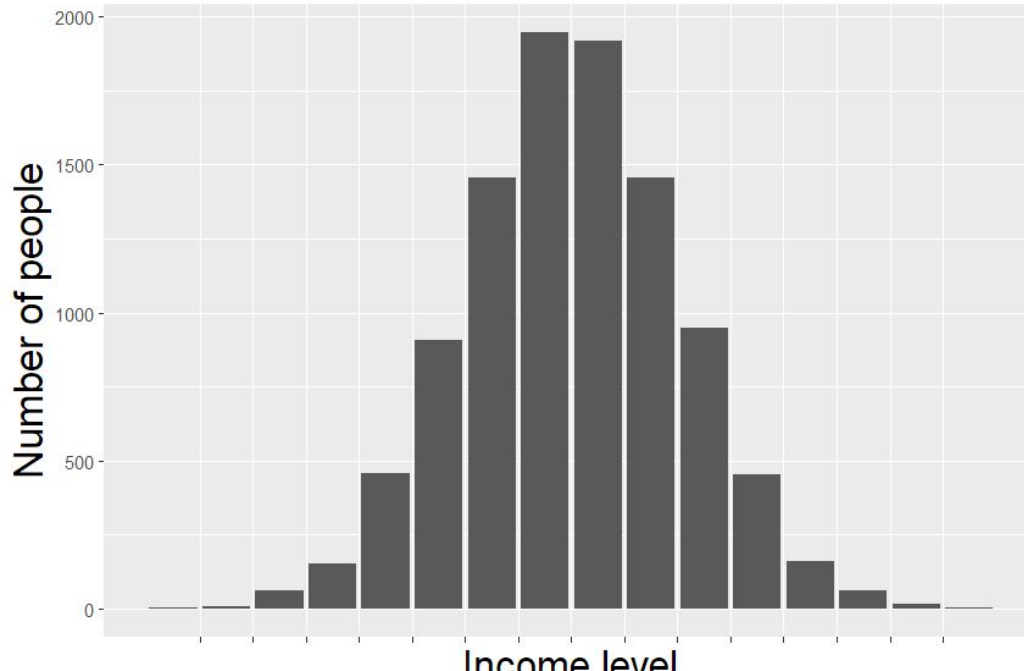*Place each letter in the corresponding grid in the table:*

	Low validity	High validity
Low reliability	D	A
High reliability	C	B

Distribution

- There are different metrics that tell us about the central tendency or "typical" value of our data:

  - Mean: The average of all values in the dataset

  - Median: The "middle" value in a dataset when all observations are ordered least to greatest

  - Mode: The most common value in a dataset

- But these metrics alone can't tell us everything. Always look at how your data is distributed - its "shape" - before making a research plan.
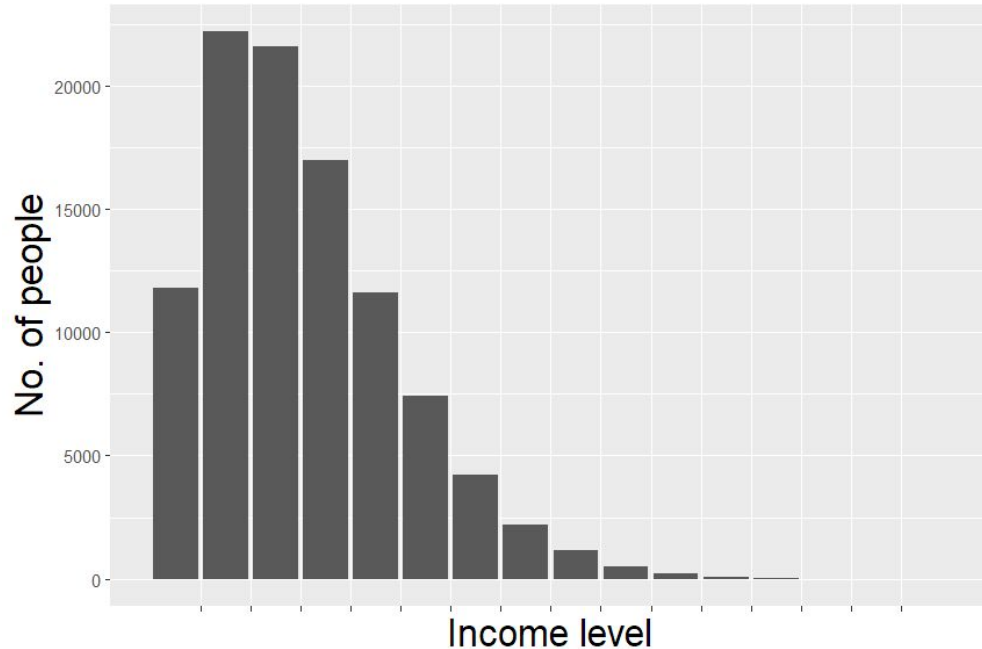
Distribution

- Always look at your data!
- Let's say you wanted to know if countries with higher income levels also had higher voter turnout. How would you measure income level in a country with an income distribution that looks like this?
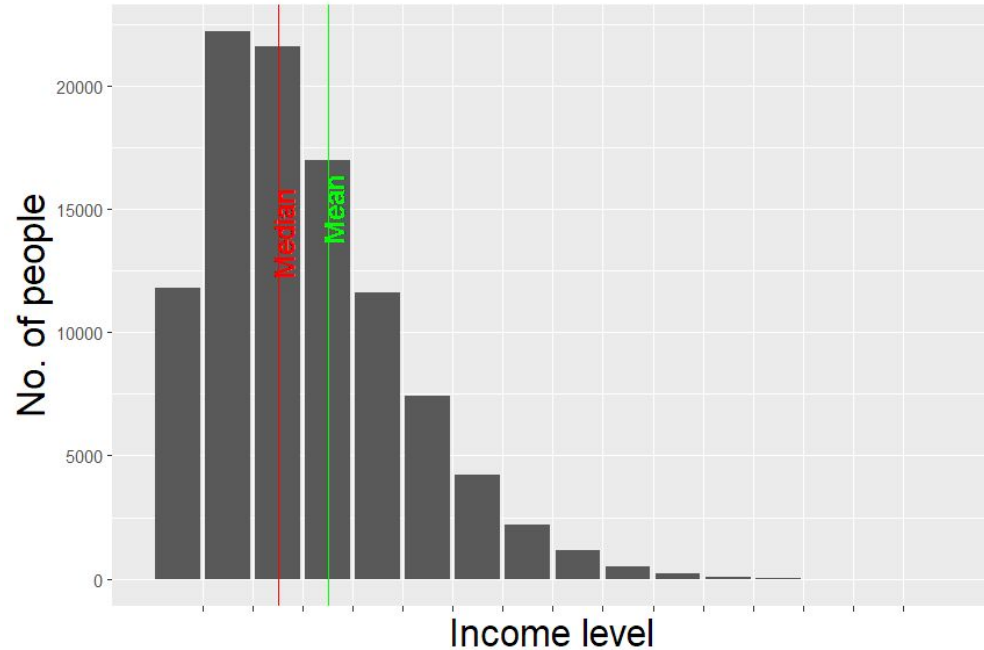
Distribution

- Always look at your data!
- Let's say you wanted to know if countries with higher income levels also had higher voter turnout. How would you measure income level in a country with an income distribution that looks like this? Or this?

Distribution

- Always look at your data!
- Let's say you wanted to know if countries with higher income levels also had higher voter turnout. How would you measure income level in a country with an income distribution that looks like this? Or this?

Distribution

- Always look at your data!
- Let's say you wanted to know if countries with higher income levels also had higher voter turnout. How would you measure income level in a country with an income distribution that looks like this? Or this? Or this?