Research Methodology for Public Policy

Lecture 4 — Measurement 1: Concepts and Variables

Dr. Colin Kuehl

2026-07-09

Northern Illinois University

SCHOOL OF PUBLIC POLICY
CHIANG MAI UNIVERSITY

Welcome Back & Warm-Up

Warm-up

Take one yellow, one blue, one green, and one pink sticky note.

1. Yellow - What concept(s) or idea(s) from our morning session do you still find confusing or would like to review one more time?
2. Blue - What are you struggling with in R?
3. Green - As of this morning, what do you think your final project will be about? Use a $\rightarrow$ to show the relationship.

ie candy -> happy students

1. Pink - What is the most interesting or important thing you've learned thus far?

Quick Recap: Lecture 3

- A good research question grows out of a gap, puzzle, or unresolved debate in the literature
- The literature review tells you what's already known — and where your project can add value
- A literature review also provides theory from which derive a hypothesis
- A good hypothesis is falsifiable, directional, and parsimonious
- Today we move from *asking* a good question to *measuring* the things the question is about

Today's Agenda

- From concepts to measurable variables
- What makes a concept well-defined
- Unit of analysis & multidimensional concepts
- Operationalizing a concept
- Types of variables and levels of measurement
- Types of data
- Describing a single variable: central tendency
- Worked example: political trust, start to finish

From Concepts to Measures

Context

- So far we've thought a lot about relationships. Today we focus on a single variable
- What is a **variable**?

Why Measurement Is Its Own Lecture

- Every number in a dataset is a stand-in for some underlying **concept**
- Good description — and good inference — depends on measuring that concept accurately
- Some contested concepts in public policy: "good governance," "state capacity," "policy success," "trust"
- "Defining a concept is one of the most essential steps in the research process."

The Measurement Chain

Concept $\rightarrow$ Operationalization $\rightarrow$ Variable $\rightarrow$ Data

- **Concept**: the abstract idea you actually care about (e.g., "political trust", "gender equality", "sustainable behavior")
- **Operationalization**: the concrete procedure that turns the concept into something you can observe
- **Variable**: the specific measured quantity that results, with values that vary across cases
- **Data**: the collected values of that variable for your actual units

The Measurement Chain

Concept $\rightarrow$ Operationalization $\rightarrow$ Variable $\rightarrow$ Data

- **Concept**: *Support for Global Environmental Cooperation* is the extent to which individuals want their country to engage in formal treaties to prevent environmental harm.
- **Operationalization**: Survey question: Do you think [your country] should do more, less, or about the same in their efforts to support the Paris climate agreement?
- **Variable**: values range from 1-5
- **Data**:

Why This Matters

- If your measure doesn't capture your concept, nothing downstream can save the analysis
- A perfectly-run regression on a badly-measured variable still produces a wrong answer
- "Garbage in, garbage out" — measurement is where most research projects quietly go wrong
- Authoritarianism, corruption, populism, civil war: common political-science terms with no single agreed-upon definition

Defining Concepts Well

Key Properties of a Good Concept

- Must be **concrete** enough to recognize when you see it
- Must **vary** — if it doesn't vary across your cases, you can't explain anything with it
- Identifying a new concept and describing its variation is a great way to start a research project

What a Conceptual Definition Needs

A conceptual definition should:

- Specify the **unit of analysis** the concept applies to
- Suggest the **characteristics** that constitute the concept
- Provide a theoretical way of telling *how much* of the characteristic a unit has — the concept must be measurable, at least in principle

A Template for Conceptual Definitions

The concept of ***[CONCEPT]*** *is defined as the extent to which* ***UNIT OF ANALYSIS*** *exhibit the characteristic of* ***[CONSTITUTIVE CHARACTERISTICS]****.*

- e.g., "Liberalism is the extent to which individuals support increased government spending on social programs"
- Try this template with a concept from your own research interest

Why Definitions Are So Contested

- "What does it mean for a country to have 'low levels of political freedom'?" — does that include freedom of speech? Freedom of the press? Freedom to compete in elections?
- Different researchers reasonably define the same concept differently depending on their research question
- This is normal — the goal is a *transparent, defensible* definition, not the one true definition

Check-in:

Check-in:

With a partner, come up with a conceptual definition of literacy.

Then, provide an operationalization for measuring literacy in a country.

Unit of Analysis & Multidimensional Concepts

Unit of Analysis

- A **unit of analysis** is that for which information is being collected
- Individuals? Households? Legislators? Bills? Countries? Trade agreements? - usually our rows in R
- Most surveys are at the individual level; household surveys mix individual- and household-level items
- Mismatches cause real confusion — be explicit about your unit

Unit of Analysis vs. Level of Analysis

- **Unit of analysis**: where the information was actually collected
- **Level of analysis**: the level at which you choose to *present* or *aggregate* the data
- Example: survey data is collected from individuals (unit), but often reported by region, income group, or country (level)
- **Ecological inference fallacy**: drawing conclusions about individuals from aggregate-level data
  - e.g., "states with more foreign-born residents have higher literacy rates" tells you nothing about whether immigrants themselves are more literate

Multidimensional Concepts

- A **conceptual dimension** is a set of concrete traits of a similar type
- A **multidimensional concept** has more than one dimension
- e.g., "liberalism" in American politics often has both an *economic* dimension and a *social* dimension
- e.g., measures of democracy combine political rights, participation, free speech, civil liberties, and competitive elections *(EMPS Ch. 4, paraphrased)*
- What other concepts in your field have multiple dimensions?

Operationalizing Concepts

From Definition to Instrument

- An **operational definition** describes the actual instrument used to measure the concept — it puts the conceptual definition "into operation"
- The conceptual definition should point the way toward the operational one
- e.g., "liberalism" operationalized as a count of how often a survey respondent favored increased government spending across a battery of items

A Concept Is Never Self-Measuring

- "The measurement of the concept involves deciding what information must be collected and how that information will be interpreted."
- For "political freedom": do we count jailed journalists? Constitutional speech guarantees? Both, weighted how?
- Every one of those choices is a judgment call — and every judgment call should be stated up front, not buried.
- We need to be transparent with these decisions. We do so in the data section of our paper

Example: Measuring Democracy

- At least 10 different published measures of democracy exist, differing in which attributes they include
- The Economist Intelligence Unit's Democracy Index: 0–10 scale, then classified into Authoritarian (under 4), Hybrid (4–6), Flawed Democracy (6–8), Full Democracy (8–10)
- The Polity score: –10 to 10, then classified into Autocracy, Anocracy, and Democracy
- Same underlying concept — different operationalizations, different aggregation rules, different conclusions

Example: Measuring Democracy

Economist Democracy Index

- Five categories: electoral process and pluralism, functioning of government, political participation, political culture, and civil liberties
- Built from 60 indicators scored largely by expert judgment, supplemented by public opinion surveys (e.g., World Values Survey) — the inclusion of citizen attitudes ("political culture") is distinctive
- Produces a 0–10 score and classifies regimes into four types: full democracy, flawed democracy, hybrid regime, authoritarian
- Treats popular support for democracy and government effectiveness as part of democracy itself, which critics argue conflates democracy with its correlates

Example: Measuring Democracy

Polity (Polity IV/5)

- Narrowest and most institutional: focuses on executive recruitment (how leaders are chosen), constraints on executive power, and political competition/participation
- Codes a Democracy score (0–10) and an Autocracy score (0–10), combined into the familiar -10 to $+10$ Polity score
- Notably excludes civil liberties, media freedom, and actual citizen participation — a country can score well with competitive elite institutions alone

Example: Measuring Democracy

V-Dem (Varieties of Democracy)

- Explicitly disaggregated: rather than one definition, it measures five separate high-level conceptions of democracy — electoral, liberal, participatory, deliberative, and egalitarian — each as its own index
- Built from hundreds of granular indicators coded by multiple country experts, aggregated with a Bayesian measurement model that estimates coder reliability and provides uncertainty estimates
- The electoral democracy (polyarchy) index is the foundation, following Dahl: suffrage, elected officials, clean elections, freedom of expression, and freedom of association; the other four indices layer additional components on top

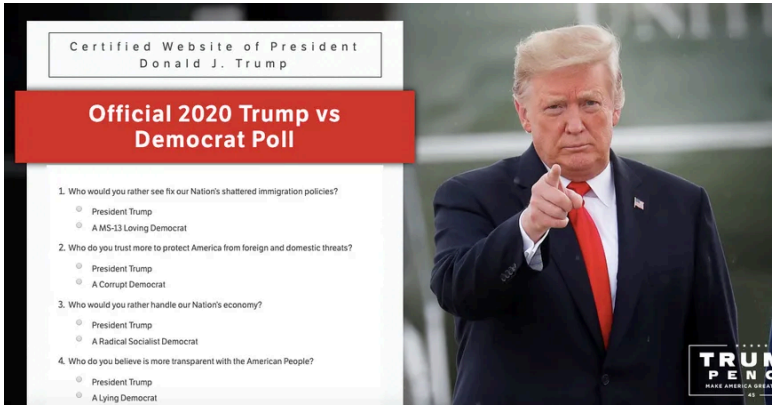
A Cautionary Tale: Measuring "Transition to Democracy"

- Researcher Pamela Paxton studied when countries "became democratic" by checking when authors coded a transition
- Most authors defined democratization as requiring universal suffrage — but routinely coded countries as democratic decades before women could vote
- Switzerland: women couldn't vote until 1971, yet most authors' own coding called it a democracy since the 19th century
- The lesson: your operational definition has to actually match the conceptual definition you wrote down

Measurement Error

- **Systematic error (bias)**: a consistent, repeated distortion — e.g., a poorly worded survey question that always pushes responses one direction, or a measuring tool that's miscalibrated
- "Systematic bias occurs when we are consistently making a mistake in our measurement." *(EMPS Ch. 4)*
- **Random error**: haphazard, inconsistent distortion — noise that should average out to roughly zero
- Re-measuring changes random error but **not** systematic error
- Larger samples reduce random error's effect on your estimates, but cannot fix systematic bias

A Real-World Example of Systematic Bias



Reliability and Validity, Briefly

- **Reliability**: every time you measure, you get the same result *(EMPS Ch. 4, paraphrased)*
- **Validity**: the measure actually represents the concept you care about *(EMPS Ch. 4, paraphrased)*
- A measure can be reliable without being valid — consistently capturing the *wrong* thing
- We dedicate all of Lecture 5 to this distinction; today, just know the terms exist

Types of Variables

The Measurement Ladder

- **Quantitative variables**: represented by numbers, even if the underlying concept isn't inherently numeric *(EMPS Ch. 4, paraphrased)*
- **Qualitative variables**: things like speech transcripts or historical documents that can't go into a dataset without substantial transformation or coding

Levels of measurement

- Within quantitative variables, four levels of measurement:

1. **Nominal / categorical**: discrete groups, no order (e.g., country, regime type, party identification)
2. **Ordinal**: ordered, but unequal spacing between categories (e.g., "strongly disagree" to "strongly agree")
3. **Interval**: ordered, equal spacing, no true zero (e.g., temperature in Celsius)
4. **Ratio**: ordered, equal spacing, *and* a true zero (e.g., income, age, vote counts)

Nominal vs Ordinal

- "Nominal variables are those that hold categories that cannot be hierarchically ordered in any way."
  - colors, race
- "Ordinal variables... can be ordered in a sequence yet we cannot establish an exact distance between them." *(EMPS Ch. 4)*
  - A survey respondent who says "strongly disagree" clearly disagrees more than one who says "agree" — but we cannot say *how much* more

Two Practical Categories, Illustrated

Continuous and Discrete Numerical Variables

- **Continuous** variables can take on any value (e.g., GDP, household income) *(EMPS Ch. 4, paraphrased)*
- **Discrete** (count) variables are integers only — there cannot be 155.3 votes in the House *(EMPS Ch. 4, paraphrased)*
- Numerical variables let us establish both order *and* exact distance between observations — something nominal and ordinal variables cannot do

Continuous vs. Discrete, Illustrated

Why the Level of Measurement Matters

- "Generally speaking, numerical data gives us more room for statistical analysis. With categorical data, the descriptive analysis is more limited." *(EMPS Ch. 4)*
- Practical advice: collect data at the *most detailed level possible* — you can always aggregate up later, but you can't disaggregate after the fact
- Example: collecting exact age in years lets you analyze a fine-grained age effect *and* still group people into generations later — collecting only the generation label forecloses that option
- In R: nominal/ordinal data are usually stored as **factors** or **strings**; interval/ratio data as **numeric** or **integer**

Check-In: Age as a Variable

- "How could age be operationalized (turned into) as different types of variables?" *(EMPS Ch. 4)*
- Discrete/ratio: exact age in years
- Ordinal: age bracket (under 18; 18–50; over 50)
- The *same* underlying concept — age — can become different types of variables depending on how you ask the question

Types of Data

Cross-Sectional, Time Series, and Panel Data

- **Cross-sectional data**: many units compared at a single point in time *(EMPS Ch. 4, paraphrased)*
- **Time series data**: repeated measures over time, but not necessarily of the same units (e.g., a quarterly poll with different respondents each wave) *(EMPS Ch. 4, paraphrased)*
- **Panel data**: the *same* units followed repeatedly over time (e.g., the British Household Panel Survey tracking the same ~10,000 individuals since 1991) *(EMPS Ch. 4, paraphrased)*
- Panel data lets you study change within a unit over time — cross-sectional data cannot
- We will focus on cross-sectional - panel and time-series require an advanced statistical knowledge

Sampling from a Population

Samples and Populations

Population	The overall collection of individuals, beyond just the sample
Sample	The collection of individuals on which statistical analyses are performed, and from which general trends for the *population* are inferred
Individual	Also called an 'object' or 'unit,' a single data point contributing to the sample

- **Population**: the full set of units you ultimately want to learn about
- **Sample**: the subset you actually observe and measure
- **Individual / unit**: a single data point within the sample
- Good sampling — covered in detail later this course — is what lets us generalize from sample back to population

Describing a Single Variable: Central Tendency

Description vs. Inference

- Our usual long-run goal is **inference**: a generalizable relationship, $X \rightarrow Y$
- But **description** matters in its own right: introducing a new dataset, polling, characterizing the state of the world
- "Descriptive statistics only allow researchers to speak about the data at hand" — generalizing beyond it is the separate job of inferential statistics
- Even when inference is the goal, you must first describe your variables accurately — this is where measurement and analysis meet

Measures of Central Tendency

Describing the "middle" of a variable:

- **Mean**: sum of all values divided by the number of values — "the most widely used summary statistic," but sensitive to outliers *(EMPS Ch. 4, paraphrased)*
- **Median**: the value that splits an ordered distribution exactly in half — far less sensitive to outliers
- **Mode**: the most frequently occurring value — usable with any variable type, especially useful for categorical data

```
# illustrative only
mean(data$income); median(data$income); table(data$region)
```

Why the Mean Can Mislead

- EMPS's example: ten seminar students earn between \$1,000–\$2,600/month; a guest speaker earns \$60,000/month
- The room's **mean** income: \$7,054 — yet all ten students earn under \$3,000
- The room's **median** income: \$2,300 — a far better description of the "typical" person in the room
- One extreme value (an outlier) can drag the mean far from where most of the data actually sits (aka the Elon Musk walks into a bar problem)

Choosing the Right Measure

- Nominal data: mode only — mean and median aren't meaningful
- Ordinal data: median or mode
- Interval / ratio data: mean, median, *and* mode are all available — choose based on skew and outliers

Worked Example: Political Trust

Step 1 — The Concept

- **Concept**: political trust — citizens' confidence that government institutions will act in their interest
- Apply the template: political trust is the extent to which **individuals** exhibit confidence in the **honesty and competence of government institutions**
- Unit of analysis: the individual citizen
- Dimensions? Trust in *which* institution — the executive, the legislature, the courts, the police? Each could be its own dimension

Step 2 — The Operational Definition

- Pick an instrument: a survey item such as "How much do you trust the national government to do what is right?"
- Response scale: "Never," "Rarely," "Sometimes," "Most of the time," "Always"
- This is an **ordinal** variable — ordered, but the gap between "Never" and "Rarely" isn't necessarily the same size as between "Sometimes" and "Most of the time"

Step 3 — Where Error Can Creep In

- **Systematic error**: a leading question wording that nudges every respondent toward "distrust," regardless of their true views
- **Random error**: a respondent has a bad day and answers more cynically than usual — pure noise
- **Validity risk**: does "trust the national government" really capture *political* trust, or partly just current-events mood?
- **Reliability risk**: would the same respondent answer the same way if surveyed again next week, all else equal?
- A self-styled "poll" with leading, loaded response options on each item
- This isn't measuring opinion — it's manufacturing a particular answer
- Notice this is a **systematic** problem: *every* respondent's answer is biased in the

Step 4 — Describing the Resulting Variable

- Once collected, you'd report: mode (the most common response category) and median (the middle category when ordered)
- You would **not** report a mean directly on the raw 1–5 labels — they're ordinal, not interval, so "2.5" doesn't have a clean interpretation
- Many surveys instead average several similar items into an *index*, which behaves more like an interval variable — more on this in Lecture 5

Worked Example: Poverty as a Concept

- **Concept**: poverty — material deprivation relative to a basic standard of living
- **Operationalization**: poverty headcount ratio at \$3.20/day (2011 PPP), % of population *(EMPS Ch. 4, paraphrased)*
- This is a **ratio**-level variable: a true zero (0% in poverty), equal spacing, meaningful ratios (a 20% rate is twice a 10% rate)

Latin American Poverty Rates — A Real Dataset

- Argentina: 2.0
- Bolivia: 11.8
- Brazil: 9.6
- Chile: 1.8
- Colombia: 10.8
- Costa Rica: 2.7
- Dom. Rep.: 5.9
- Ecuador: 8.7
- El Salvador: 8.5
- Guatemala: 24.2
- Haiti: 50.8
- Honduras: 31.6
- Mexico: 11.2
- Nicaragua: 12.8
- Panama: 6.3
- Paraguay: 5.6
- Peru: 9.8
- Uruguay: 0.4

(EMPS Ch. 4 application data — poverty headcount ratio, % of population)

Central Tendency on a Real Variable

- Mean across these 18 countries: pulled upward by Haiti's 50.8% — a clear outlier
- Median: a much more typical summary of where most countries actually sit
- Try it yourself: how much does dropping Haiti change the mean? What about dropping Uruguay (the low outlier)?
- The lesson from EMPS's income example earlier applies directly here — ratio-level data with outliers is exactly when mean and median can tell different stories

Key Terms & Looking Ahead

Key Terms from Today

- Concept / operationalization
- Unit of analysis / level of analysis
- Multidimensional concept
- Measurement error (systematic / random)
- Reliability / validity (preview)
- Nominal / ordinal / interval / ratio
- Cross-sectional / time series / panel data
- Population / sample
- Central tendency
- Mean / median / mode

Before Next Session

Lecture 5: Measurement 2 — Validity and Reliability

- Read: EMPS, Ch. 4 (review, focus on the Measurement section)
- Read: QRM, Ch. 5 (*Concepts, Operationalization, and Measurement*)
- Bring the variable you sketched today — we'll stress-test it for validity and reliability

Discussion

Pick a concept from your own research interest. What is your conceptual definition? What would your operational definition be — and where might measurement error creep in?

Questions?

Thank you — see you in the Practicum this afternoon.

See You in the Practicum