STAT/SOC/CSSS 221
Statistical Concepts and Methods for the Social Sciences

Levels of Measurement

Christopher Adolph

Department of Political Science
and
Center for Statistics and the Social Sciences
University of Washington, Seattle

Aside on Notation

Statisticians use math to express concepts clearly and succinctly

Math notation is just a way to abbreviate simple concepts

But just like in language, simple concepts combine into complex ideas

So learn notation well *before* diving in to new statistics

Today's notation:

1. How statisticians write out knowns and unknowns
2. New symbols in today's lecture

Knowns and Unknows

Statistics is concerned with using *things we know* to infer *things we don't know*

Most statistical notation places a sharp distinction between these categories

Knowns and Unknows

Statistics is concerned with using *things we know* to infer *things we don't know*

Most statistical notation places a sharp distinction between these categories

Statisticians use *words* or *roman letters* to represent known quantities:

We might name the variable representing the amount of money a person reported earning as y or Income.

and the variable representing the sex of the respondent as x or Female.

Knowns and Unknows

Statistics is concerned with using *things we know* to infer *things we don't know*

Most statistical notation places a sharp distinction between these categories

Statisticians use *words* or *roman letters* to represent known quantities:

We might name the variable representing the amount of money a person reported earning as y or Income.

and the variable representing the sex of the respondent as x or Female.

Statisticians use *Greek letters* to represent unknown quantities:

We might denote the effect of being female on income (e.g., the cumulative effect of discrimination or structural disadvantage) as β

Learn the lowecase Greek alphabet!

α	`alpha`	κ	`kappa`	σ	`sigma`
β	`beta`	λ	`lambda`	τ	`tau`
γ	`gamma`	μ	`mu`	υ	`upsilon`
δ	`delta`	ν	`nu`	ϕ	`phi`
ε	`epsilon`	ξ	`xi`	ψ	`psi`
ζ	`zeta`	o	`omicron`	χ	`chi`
η	`eta`	π	`pi`	ω	`omega`
θ	`theta`	ρ	`rho`		

This won't be tested *per se*, but familiarity with these letters will greatly aid comprehension as the quarter progresses.

Today's new notation (more than usual)

∞	Infinity. Comes in positive (∞) and negative ($-\infty$) varieties.

Today's new notation (more than usual)

∞	Infinity. Comes in positive (∞) and negative ($-\infty$) varieties.
$\{a, b, c\}$	A set containing elements a, b and c.
$\in$	"is in the set": an operator establishing the element on the left is in the set on the right.

Today's new notation (more than usual)

∞	Infinity. Comes in positive (∞) and negative ($-\infty$) varieties.
$\{a, b, c\}$	A set containing elements a, b and c.
$\in$	"is in the set": an operator establishing the element on the left is in the set on the right.
$\mathbb{R}$	The set of Real numbers (every possible decimal value). This set contains an infinite number of items!

Today's new notation (more than usual)

∞	Infinity. Comes in positive (∞) and negative ($-\infty$) varieties.
$\{a, b, c\}$	A set containing elements a, b and c.
$\in$	"is in the set": an operator establishing the element on the left is in the set on the right.
$\mathbb{R}$	The set of Real numbers (every possible decimal value). This set contains an infinite number of items!
$\mapsto$	"maps to": an operator establishing a correspondence between the elements of one set and another, like an English-to-Spanish dictionary does with words.

Continuous & discrete data

All variables are either continuous or discrete

This determines which statistical tools are the right ones for your dependent variable (the variable whose pattern of variation you are trying to explain)

Discrete data can be matched up to the integers. There is a clear distinction between each possible value a discrete variable may take on.

Examples: Your sex; Number of cities you have lived in

Continuous & discrete data

All variables are either continuous or discrete

This determines which statistical tools are the right ones for your dependent variable (the variable whose pattern of variation you are trying to explain)

Discrete data can be matched up to the integers. There is a clear distinction between each possible value a discrete variable may take on.

Examples: Your sex; Number of cities you have lived in

Continuous data can take on any real value between a lower and upper bound. If the upper and lower bounds are $[-\infty, \infty]$, then a variable can take on any numerical value.

Examples: The unemployment rate; a family's net worth

Aside: Integers, Real Numbers, & Infinity

Infinity (∞) is a tricky mathematical concept, but one tied up with the distinction between discrete and continuous variables

Integers are the negative whole numbers, positive whole numbers, and zero:

$$-\infty, \ldots, -1000, -999, \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots, 999, 1000, \ldots, \infty$$

There are infinitely many integers

Aside: Integers, Real Numbers, & Infinity

Real numbers include every possible decimal within a given interval:

$$\mathbb{R} \in (\ell, u)$$

We can't list the real numbers, even using "..."

Between any two real numbers there are more real numbers.

In fact, there are an uncountable infinity of additional real numbers between *any* two real numbers

Discrete variables

There are three types of discrete variables: Binary, Ordered, & Nominal

Binary data take on only two possible values.
Without loss of generality, let these values be 0 and 1.

Examples:

Did you vote? $\{\text{No}, \text{Yes}\} \mapsto 0, 1$

Are you a Catholic? $\{\text{No}, \text{Yes}\} \mapsto 0, 1$

Discrete variables

Ordered data take on countably many values.
That is, it can be mapped to a subset of the counting numbers: $1, 2, 3, \ldots$

Examples:

Do you support 2010 Health Care reform?

$$\{\text{Does too little, Just right, Doesn't do enough}\} \quad \mapsto \quad \{1, 2, 3\}$$

Discrete variables

Ordered data take on countably many values.
That is, it can be mapped to a subset of the counting numbers: $1, 2, 3, \ldots$

Examples:

Do you support 2010 Health Care reform?

$$\{\text{Does too little, Just right, Doesn't do enough}\} \quad \mapsto \quad \{1, 2, 3\}$$

How democratic is a given country? (Polity IV)

$$\{-10, -9, \ldots, -1, 0, 1, \ldots, 9, 10\} \quad \mapsto \quad \{\text{Authoritarianism}, \ldots \text{Democracy}\}$$

Discrete variables

Ordered data take on countably many values.
That is, it can be mapped to a subset of the counting numbers: $1, 2, 3, \ldots$

Examples:

Do you support 2010 Health Care reform?

$\{\text{Does too little}, \text{Just right}, \text{Doesn't do enough}\} \quad \mapsto \quad \{1, 2, 3\}$

How democratic is a given country? (Polity IV)

$\{-10, -9, \ldots, -1, 0, 1, \ldots, 9, 10\} \quad \mapsto \quad \{\text{Authoritarianism}, \ldots \text{Democracy}\}$

How many people vote for a candidate in an election?

$\{0, 1, 2, 3 \ldots m\}$, where m is the number of registered voters

Discrete variables

Ordered data take on countably many values.
That is, it can be mapped to a subset of the counting numbers: $1, 2, 3, \ldots$

Examples:

Do you support 2010 Health Care reform?
$\{\text{Does too little, Just right, Doesn't do enough}\} \quad \mapsto \quad \{1, 2, 3\}$

How democratic is a given country? (Polity IV)
$\{-10, -9, \ldots, -1, 0, 1, \ldots, 9, 10\} \quad \mapsto \quad \{\text{Authoritarianism}, \ldots \text{Democracy}\}$

How many people vote for a candidate in an election?
$\{0, 1, 2, 3 \ldots m\}$, where m is the number of registered voters

How many times does the press mention the Iowa caucus today?
$\{0, 1, 2, 3 \ldots \infty\}$

Discrete variables

Nominal data take on name values lacking a unique ordering

Examples:

Which candidate do you prefer? $\{\text{Obama}, \text{Romney}, \text{Another} - \text{Republican}\}$

Which region do you live in? $\{\text{Northeast}, \text{Midwest}, \text{South}, \text{West}\}$

Discrete variables

We can't map the coding of Nominal variables to any ordering

But notice we can recode *any* discrete variable as a series of binary variables:

Which candidate do you prefer? $\mapsto$

1. Do you prefer Romney to Obama & Another-Republican?
   $\{\text{No}, \text{Yes}\} \mapsto 0, 1$
2. Do you prefer Obama to Romney & Another-Republican?
   $\{\text{No}, \text{Yes}\} \mapsto 0, 1$
3. Do you prefer Another-Republican to Obama & Romney?
   $\{\text{No}, \text{Yes}\} \mapsto 0, 1$

Also notice that any two of these questions is sufficient to reconstruct the full Nominal variable

Continuous variables

A continuous variable is one that can take on any *real* value

Examples:

- Unemployment rate: Can take on any real value between 0 and 100%.

Continuous variables

A continuous variable is one that can take on any *real* value

Examples:

- Unemployment rate: Can take on any real value between 0 and 100%. (Or can it? Close enough?)

Continuous variables

A continuous variable is one that can take on any *real* value

Examples:

- Unemployment rate: Can take on any real value between 0 and 100%. (Or can it? Close enough?)
- Gross domestic product: Can take on any positive real value. (Or can it? Close enough?)

Continuous variables

A continuous variable is one that can take on any *real* value

Examples:

- Unemployment rate: Can take on any real value between 0 and 100%. (Or can it? Close enough?)
- Gross domestic product: Can take on any positive real value. (Or can it? Close enough?)
- Growth in gross domestic product: Any positive real value (Close enough?)

Continuous variables

A continuous variable is one that can take on any *real* value

Examples:

- Unemployment rate: Can take on any real value between 0 and 100%. (Or can it? Close enough?)
- Gross domestic product: Can take on any positive real value. (Or can it? Close enough?)
- Growth in gross domestic product: Any positive real value (Close enough?)
- Inequality: Ratio of 90th to 10th percentile of income (Close enough?)

Notice lots of economic variables are continuous

But most social and political variables are discrete!

Additive versus Ratio scales

Continuous variables come in two flavors, depending on whether zero really means an absence of the variable:

Additive versus Ratio scales

Continuous variables come in two flavors, depending on whether zero really means an absence of the variable:

Additive No meaningful zero $\rightarrow$ “1 unit increase” has a consistent meaning across the scale, but a ratio of two levels does not.

Additive versus Ratio scales

Continuous variables come in two flavors, depending on whether zero really means an absence of the variable:

Additive No meaningful zero $\rightarrow$ "1 unit increase" has a consistent meaning across the scale, but a ratio of two levels does not.

Examples:
Degrees Fahrenheit

Additive versus Ratio scales

Continuous variables come in two flavors, depending on whether zero really means an absence of the variable:

Additive No meaningful zero $\rightarrow$ "1 unit increase" has a consistent meaning across the scale, but a ratio of two levels does not.

Examples:
Degrees Fahrenheit
Polity IV Democracy Score

Additive versus Ratio scales

Continuous variables come in two flavors, depending on whether zero really means an absence of the variable:

Additive No meaningful zero $\rightarrow$ "1 unit increase" has a consistent meaning across the scale, but a ratio of two levels does not.

Examples:
Degrees Fahrenheit
Polity IV Democracy Score
Feeling Thermometer Scores

Additive versus Ratio scales

Continuous variables come in two flavors, depending on whether zero really means an absence of the variable:

Additive No meaningful zero $\rightarrow$ "1 unit increase" has a consistent meaning across the scale, but a ratio of two levels does not.

Examples:
Degrees Fahrenheit
Polity IV Democracy Score
Feeling Thermometer Scores

Ratio Meaningful zero $\rightarrow$ "1 unit increase" has a consistent meaning across the scale, & ratio convey "factor changes"

Additive versus Ratio scales

Continuous variables come in two flavors, depending on whether zero really means an absence of the variable:

Additive No meaningful zero $\rightarrow$ “1 unit increase” has a consistent meaning across the scale, but a ratio of two levels does not.

Examples:
Degrees Fahrenheit
Polity IV Democracy Score
Feeling Thermometer Scores

Ratio Meaningful zero $\rightarrow$ “1 unit increase” has a consistent meaning across the scale, & ratio convey “factor changes”

Examples:
Degrees Above Absolute Zero

Additive versus Ratio scales

Continuous variables come in two flavors, depending on whether zero really means an absence of the variable:

Additive No meaningful zero $\rightarrow$ "1 unit increase" has a consistent meaning across the scale, but a ratio of two levels does not.

Examples:
Degrees Fahrenheit
Polity IV Democracy Score
Feeling Thermometer Scores

Ratio Meaningful zero $\rightarrow$ "1 unit increase" has a consistent meaning across the scale, & ratio convey "factor changes"

Examples:
Degrees Above Absolute Zero
Number of Votes Received

Additive versus Ratio scales

Continuous variables come in two flavors, depending on whether zero really means an absence of the variable:

Additive No meaningful zero $\rightarrow$ "1 unit increase" has a consistent meaning across the scale, but a ratio of two levels does not.

Examples:
Degrees Fahrenheit
Polity IV Democracy Score
Feeling Thermometer Scores

Ratio Meaningful zero $\rightarrow$ "1 unit increase" has a consistent meaning across the scale, & ratio convey "factor changes"

Examples:
Degrees Above Absolute Zero
Number of Votes Received
Unemployment Rate