

Foundations of Quantitative Social Science

Day 1 AM

Sarah Moore & J. Seawright

Northwestern University

Math Camp 2023

Day 1 Agenda

- Get to know each other.
- Acquaint ourselves with the foundations of quantitative social science.
- Today we will jump around a little bit more than the rest of the days to get through some necessary review:
 - Sets and basics of set operations
 - Algebra refresh

Who are we?

Co-Instructors:

J. Seawright

- Professor of Political Science and Director of Graduate Studies
- j-seawright@northwestern.edu

Sarah Moore

- PhD Candidate (ABD) in Political Science and M.S. Candidate in Statistics
- sarahmoore2022@u.northwestern.edu

What is math camp?

- A primer to (re)introduce you to some topics that will be necessary to moving forward in the quantitative methods sequence in Political Science and Sociology.
- Allow the instructors to assess the needs of the incoming cohort regarding instructional support.
- Also an opportunity to meet and network with your new grad school colleagues.

Now, tell us about you

Let's get to know each other a bit more:

- Name
- Pronouns
- Discipline + Subfield
- Favorite comfort food

What are we going to cover?

- Refresh knowledge of foundational mathematical concepts
- Establish base knowledge of computational tools used in social science research.
- Obtain familiarity with expectations of a graduate level course in methods.
- Build rapport with cohort members, upper level graduate students, and faculty.
- Think broadly about the potential use of quantitative methods in social science.
- Reflect on how well you understand and can apply the concepts addressed in math camp.

Expectations for Math Camp

- Be respectful of your fellow Math Campers and respect that everyone is here to learn!
- Review materials (text, slides, problem sets)
- Complete problem sets by the scheduled due date
- Complete R exercises
- We don't expect that you will learn everything in depth. This camp is basically just to introduce you to some concepts that you will run into throughout methods courses.

**What's the point of taking
methods courses to study
[insert discipline here] ?**

Theory Construction

Social scientists develop and test **theories** that explain certain patterns that exist in the social world.

These theories emerge out of research questions related to theoretical or empirical puzzles or problem-based inquiries.

Theory

A **theory** is a proposition or set of propositions that explain a certain phenomenon. Theories can be either *causal* or *descriptive*, though more work in contemporary social science is now oriented toward causal explanation.

A theory specifies a number of **concepts** of interest and explains the relationship between them.

Observable Implications of a Theory

Empirical social scientists evaluate whether or not theories are true through hypothesis testing. Hypothesis testing is a systematic method of discerning if observed evidence favors the theory given what we expect to observe if the theory were true.

Hypothesis and Null Hypothesis

A **hypothesis** is an observation we would expect if a theory is true.

A **null hypothesis** is the corresponding observation we would expect if the theory is *not* true.

Variables

To test the observable implications of theory, we have to systematize how we'd expect to observe it empirically. This first entails translating theoretically abstract concepts into empirical referents that represent concepts--- these are *variables*.

- The **dependent variable** (Y) is the outcome of interest, whose value is expected to depend on some other variable. The dependent variable may also be referred to as the outcome or response variable.
- The **independent variable** (X) is the theorized cause of the dependent variable, or at least the cause of Y's variation.

Operationalization

For example, we might be abstractly interested in *socioeconomic inequality*. How do you all suggest that we turn this concept into a variable?

There's no right answer here...There are **many** ways to measure this concept.

When we think of **operationalizing** socioeconomic inequality, or any other concept of interest, we need to choose a fixed, systematic measure of this concept to help us identify instances of socioeconomic inequality.

Developing Hypotheses, An Example

Our theory states that public approval of politicians is grounded in more than their *present* political careers. We believe that politician approval is dependent on their social activity and image preceding their political careers and that pre-term high profile politicians enjoy more popular approval.

- What are some concepts in this theory and how might we operationalize them into variables?
- What is a hypothesis that we can derive from this theory? That is, **what would we observe if this theory is true?**

Theory Testing

We need tools to evaluate our hypotheses relative to data. A theory's contribution to knowledge and understanding of the world depends on how well the theory performs under these hypothesis tests.

- Empirical methods can be either *qualitative* or *quantitative* in nature-- here we focus on the latter.
- Quantitative methods use statistical or mathematical analyses to determine whether the accumulated evidence favors the hypothesis proposed relative to the null. Our default is to *favor the null*, and the burden is on the data to show otherwise.
 - There are more nuances to hypothesis testing that we will get into later next week.

From Theory to Empirics

- Observations about the world around us constitute *data*.
- When we think about observing quantitative data, there are a lot of things to keep in mind: the bounds, the set of possible values, quantifying abstract concepts, and more.
- While we will use applied probability and statistics to reach conclusions about our data, there are intermediate steps and additional keys to navigating the inter-related processes of data collection and analysis.
- These intermediate steps are based in logic, set theory, algebra, calculus, etc. We will introduce or re-introduce all of these constitutive parts so that once we get to the coursework in 403 and beyond, you can always revisit this material to make sense of the foundations of more advanced methods.

Sets and Logical Possibilities

- We measure and assign values to higher order concepts that we turn into variables. How we measure something and the bounds of its measurement are consequential to how we then analyze it.

Sets

A collection of elements.

You can have a set composed of numbers or other *things*, such as names of people, places, or other non-numeric values.

$A = \{\text{cats, rabbits, dogs}\}$

$B = \{0.5, 0.15, 0.35\}$

Types of numerical elements

- Natural numbers, \mathbb{N} : positive whole numbers, including 0: $\{(0), 1, 2, 3, \dots\}$
- Integers, \mathbb{Z} : negative and positive numbers including zero: $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
- Rational numbers, \mathbb{Q} : any number that can be represented by the quotient p/q : $\{1/2, 4, -3, 3/2\}$
- Real numbers, \mathbb{R} : any point on the number line: $\{-899.8, 22, 4/5, \sqrt{\pi}\}$
- Imaginary numbers, \mathbb{I} : a number times the imaginary number i , which has no meaningful value, $i = \sqrt{-1}$
- Complex numbers, \mathbb{C} : a value consisting of both a real and imaginary number, often in the form $a + bi$

Measurement of elements in a set

Qualitative Difference

- **Nominal**: no mathematical relationship between the variables, values are labels with no ordered relationship
- **Ordinal**: categorical variable with set, non-numeric relationship between values, order of values is meaningful

Quantitative Difference

- **Interval**: equidistant from one value to the next, but there is no meaningful zero
- **Ratio**: equidistant from one value to the next, there is a meaningful zero

Measurement: Examples

Here are some examples of each of these types of measurements:

- **Nominal**: Classes available in the sociology department
 - Stratification, Social Networks, Collective Memory
- **Ordinal**: Your satisfaction with classes you have taken in grad school on a scale from 1 to 4
 - Very Dissatisfied (1) - (2) - (3) - Very Satisfied (4)
- **Interval**: Scores on entrance exams
 - 130-170 on GRE
- **Ratio**: The number of hours that you spend studying for classes each week
 - 0 hours, 6 hours, 18 hours

What are some other examples for these measurement types?

Sets, Variables, and Sample Spaces

- Once we know *how* to characterize variables relative to their higher order concept, must also consider:
 - Sample space: set containing all the values a variable can take
 - Subsets: subgroups of a set or sample space

Set Notation

Sets and set operations have notation conventions:

- There exists, \exists
- For all, \forall
- Union \cup ; Intersection \cap
- Excluding \notin , \neg
- Empty set \emptyset
- Element \in
- Equivalent \equiv
- Such that (s.t) or $|$, e.g. $\{x|x > 7\}$
- Subset \subset , \subseteq (these function roughly like the less than/less than equal to, but for sets)

Displaying Sets

- Roster notation: $\{x_1, x_2, \dots\}$, simply lists out the elements of the set
- Set builder notation: $\{x|x \in N, x < 4\}$, uses set notation to specify the parameters of the set
 - Try one: $A = \{x|x \in \mathbb{R}, x < 200\}$
 - Read "Set A is a set of all x such that x is a real number less than 200."

Special Kinds of Sets

- **Empty Set**: no elements.
- **Universal Set**: all elements
- **Open sets** have a fuzzy boundary. This is the set version of open brackets ().
Closed sets have clear and firm boundary. This is the set version of closed brackets []

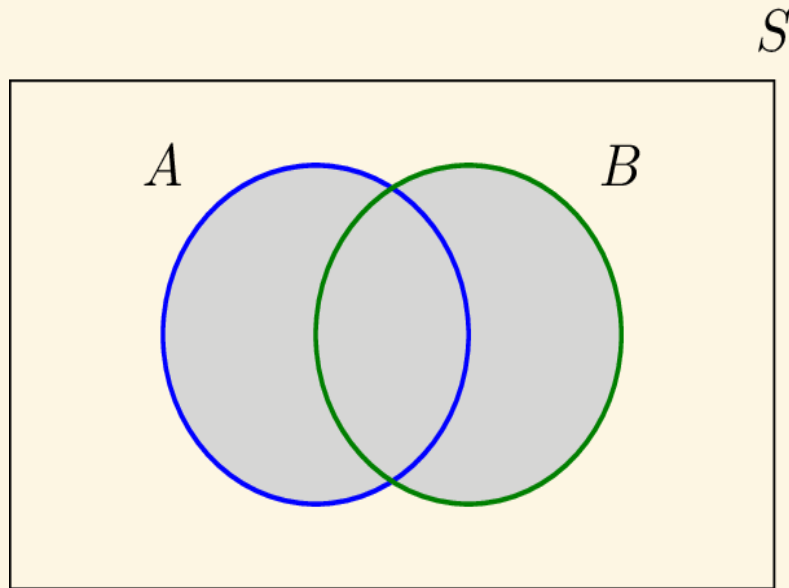
Set Operations

The images on the following slides were pulled from [this site](#).

Union: $A \cup B$

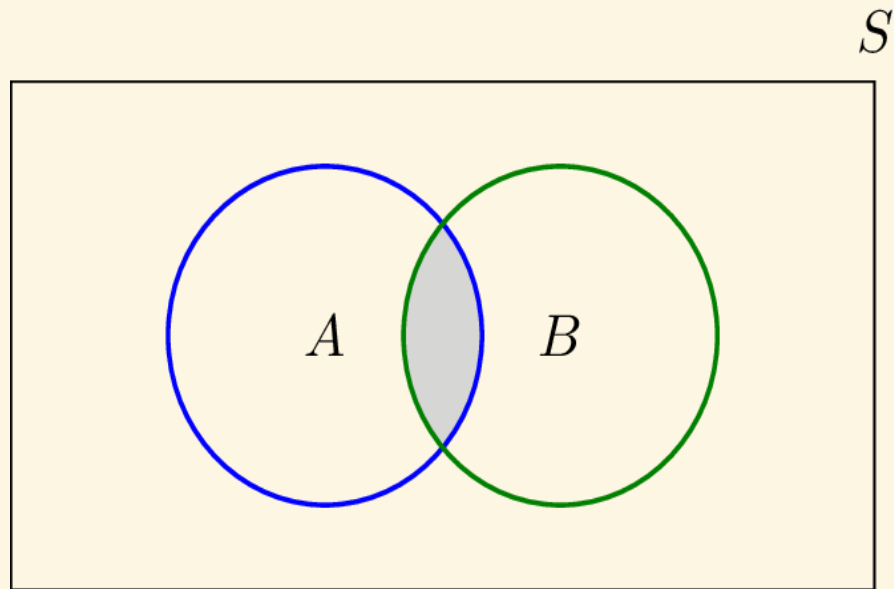
Joining two sets together, the equivalent of adding all of the elements of each set.

Here we highlight all of the elements in A OR B .



Intersection: $A \cap B$

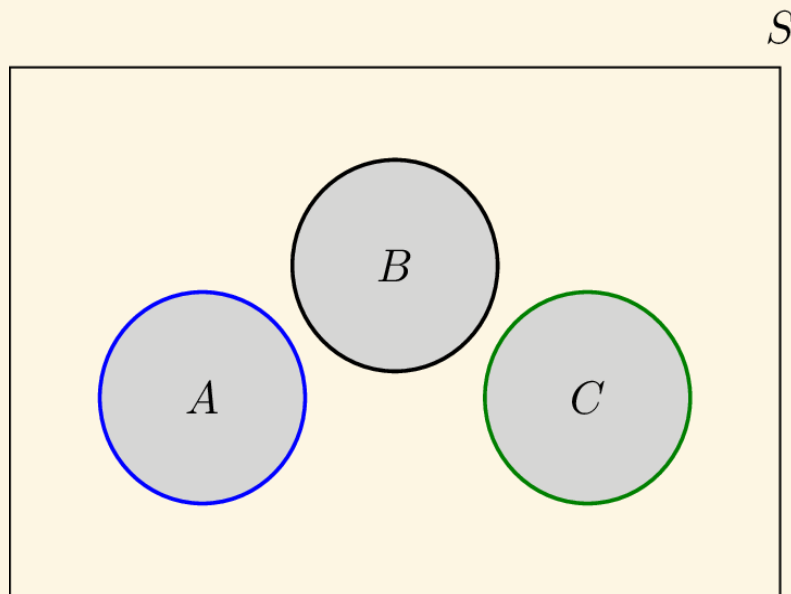
Finding the elements that are shared between 2+ sets. Here we highlight all of the elements in A AND B .



Disjoint Sets:

- Two or more sets are **disjoint** if their intersection is empty.

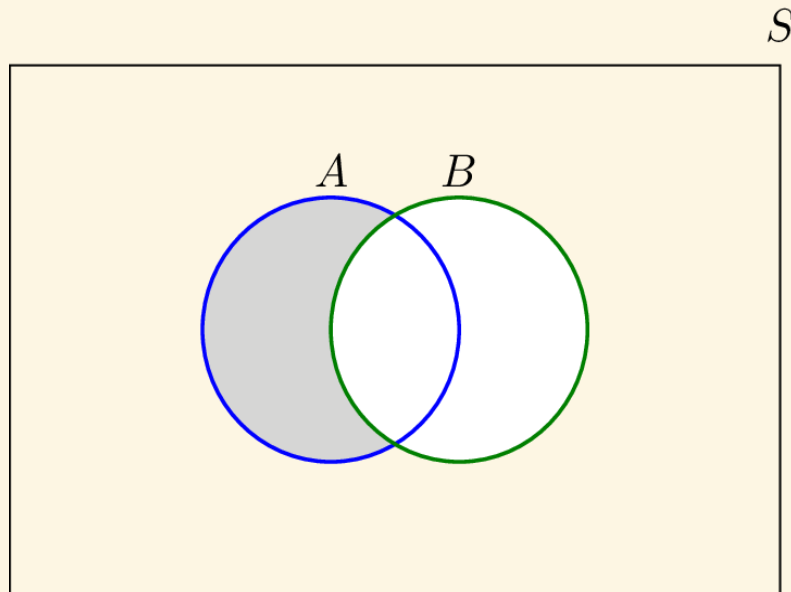
For example, $A \cap B \cap C = \emptyset$.



Difference: $A - B$

The elements of one set less the elements of another.

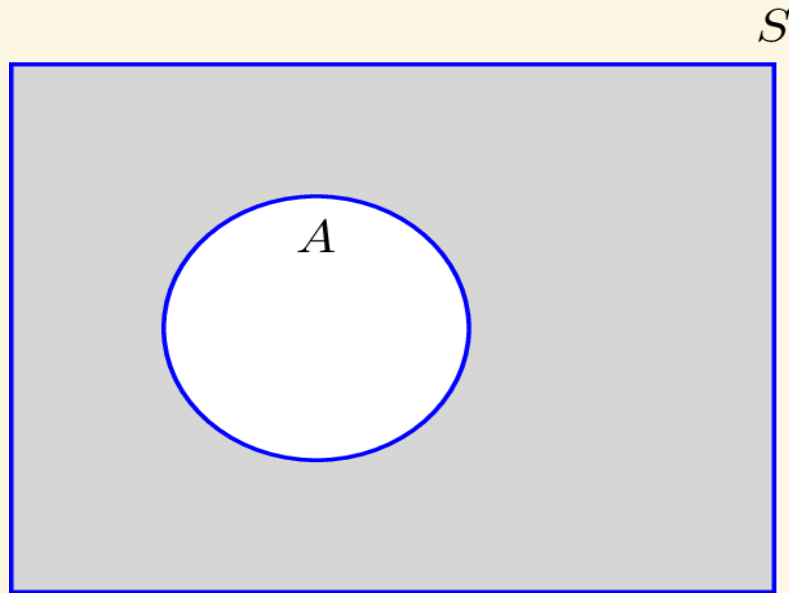
Here we show the difference $(A - B)$.



Complement: A^C

The elements in the universal set U not in the specified set.

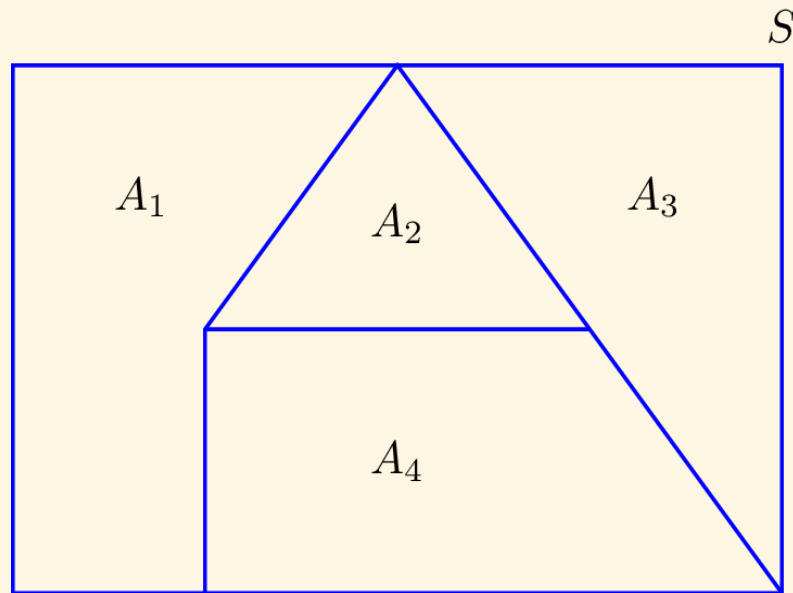
For example, A^C indicates all the elements *NOT* in A .



Partition

A grouping of elements in a universal set such that each element belongs to at least one subset.

Here all of the elements of the universal set S are grouped into subsets A_1, A_2, A_3, A_4 .



Cartesian Product: $A \times B$

Set of all the ordered pairs possible between the elements of A and B .

```
a = c('trees', 'apples', 'rivers')  
b = c('pencil', 'paper', 'chalk')  
  
expand.grid(a = a, b = b)
```

```
##      a      b  
## 1 trees pencil  
## 2 apples pencil  
## 3 rivers pencil  
## 4 trees  paper  
## 5 apples paper  
## 6 rivers paper  
## 7 trees  chalk  
## 8 apples chalk  
## 9 rivers chalk
```

Logical and Relational Operators

notation	meaning
&	AND
	OR
!	NOT
<	LESS THAN
≤	LESS THAN OR EQUAL TO
>	GREATER THAN
≥	GREATER THAN OR EQUAL TO
==	EXACTLY EQUAL TO
!=	NOT EQUAL TO
%in%	IN THE SET

Algebraic Properties

1. **Associative property** $(a + b) + c = a + (b + c)$ and $(a \times b) \times c = a \times (b \times c)$
2. **Commutative property** $a + b = b + a$ and $a \times b = b \times a$
3. **Distributive property** $a(b + c) = ab + ac$
4. **Identity property** $x + 0 = x$ and $x \times 1 = x$
5. **Inverse property** $-x + x = 0$.

Multiplicative inverse exists, but not for all numbers $x^{-1} \times x = 1$

Order of Operations

PEMDAS:

1) Parentheses

2) Exponents

3a) Multiplication (left-to-right)

3b) Division(left-to-right)

4a) Addition (left-to-right)

4b) Subtraction (left-to-right)

$$(3^3 \times 2) + 2 \times 3$$

$$= 60$$

Intervals and Inequalities

Ranges of numbers can be expressed with either $[]$ or $()$ brackets.

Notation	Meaning	Explanation
$[a, b]$	$a \leq x \leq b$	Inclusive of end points (closed interval)
(a, b)	$a < x < b$	Exclusive of end points (open interval)
$\{a, b\}$		Typically used for sets -- not inequalities/intervals

Relational Expressions

- $4 < x$
- $y > 12$
- $3 < z < 7$
- $(3, 9)$
- $[-7, 2)$

Absolute Value

Magnitude of a numerical value indicated by $|x|$. Solved by typical arithmetic, but must solve for both the negative and positive equations.

For example:

$$|x + 4| < 5$$

- Must be solved as: $-x - 4 < 5$ AND $x + 4 < 5$

Therefore:

- $-9 < x < 1$ Notice how the inequality was flipped due to the negative equation
- Try this one out: $|\frac{x+2}{x}| < 10$
- Answer: $x < -\frac{2}{11}$ or $x > \frac{2}{9}$

Exponents

- Multiplying a number by itself: $x * x * x * x = x^4$
- Add or subtract exponents when multiplying same variable: $x^4 * x^{-3} = x$
- Simplify bases when same exponent: $x^4 * y^4 = (xy)^4$ and $m^{-2}n^2 = (m^{-1}n)^2$
- Power of power: multiply, $(x^3)^3 = x^{3*3} = x^9$

Try: $xz^2(x^3z^{-2})^3$

$$= xz^2(x^9z^{-6})$$

$$= x^{10}z^{-4}$$

Logarithms

Logarithms, or logs, are the inverses of exponents: the power to which you raise the base, e.g. 10, to produce a given value, e.g. \$z\$

Logarithms (typically base 10 ($\log(x)$) or base e ($\ln(x)$), but any base is possible, e.g. $\log_{8675309}x$ (Bases aside from e and **10** will be specified explicitly).

- $y = \log(z) \leftrightarrow 10^y = z$
- $y = \ln(z) \leftrightarrow e^y = z$
- $\log(1) = 0$

Logs: Simplify the following

$$\log(x^4)$$

$$\log(xy)$$

$$\ln(e^3)$$

$$\ln(1)$$

$$\log(3) + \log(7)$$

Logs: Simplify the following

$$\log(x^4) = 4\log(x)$$

$$\log(xy) = \log(x) + \log(y)$$

$$\ln(e^3) = 3$$

$$\ln(1) = 0$$

$$\log(3) + \log(7) = \log(21)$$

$$\log(x/y) = \log(x) - \log(y) \text{ provided } (x,y > 0)$$

Factoring

We may need to break down different functions.

$$x^2 + 3x - 4 = (x - 1)(x + 4)$$

$$2x^2 + 7x + 3 = (2x + 1)(x + 3)$$

Summation

Using sums can help us when we're summing a large number of items or when summing more complex functions. Here's the general notation:

$$\sum_{i=1}^N x_1 + x_2 + \cdots + x_n$$

Σ : Sum

N : Upper limit of summation

i : Element index

Product

We can apply the same idea to an operator indicating the product of terms:

$$\prod_{i=1}^N x_1 \times x_2 \times \cdots \times x_n$$

Sequences and Series

Sequences

A sequence is an ordered list of numbers. They can be infinite or finite, but all are *countable*. Written $x_{i=1}^N$

Series

A series is the sum of a sequence, written $\sum_{i=1}^N x_i$.

Sequences

- We refer to the elements of a sequence by their position. For example, the third element would be x_3 .
- A sequence is generated and represented by some equation wherein each element is a function of the element preceding it.
- Take the following sequence: $\{1, 2, 5, 26, 677...\}$. This sequence is defined by the following generating equation:

$$x_i^2 + 1_{i=1}^{\infty}$$

- Some sequences are *infinite*, as the sequence above. Others are *finite* and will end at the n th term.

Series

- A series will essentially take a sequence and add together each of the numbers, such that our example sequence from the previous slide becomes: $\{1 + 2 + 5 + 26 + 677...\}$.
- Infinite series will sum to ∞ given that they are infinite! Shocking! However, we might also specify to know the sum at a given n term. In which case I would probably just tell you to look up a calculator to do it for you.
- Finite sums can more easily be summed, again likely with a calculator if N is sufficiently large.

If we have time...

Software!!!!