

Day 1: Foundations of Quantitative Social Science

Sarah Moore and J. Seawright

Math Camp 2023

Day 1 Agenda

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 and/or intended research area
- ▶ A hobby or interesting fact about yourself

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

- ▶ Come to every session, we will be on time
- ▶ Review materials (text, slides, problem sets)
- ▶ Complete problem sets by the scheduled due date
- ▶ Complete R exercises

Broad Expectations as a Grad Student

- ▶ Read what's assigned, and learn how to skim now.
- ▶ The people around you are your colleagues, not your competition. Our default here is to treat everyone with respect.
- ▶ Be the scholar *you* want to be, not the scholar you think everyone else wants you to be.

Theory Construction

Social scientists are tasked with developing and testing **theories** that explain certain patterns that exist in the social world.

Theory

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

Typically, a theory specifies a number of **concepts** of interest and explains the relationship between them. The way social scientists define theoretical concepts depends on previous scientific work on that phenomenon and their manifestation in real life.

Theory Construction

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 developed theory or not.

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.

A hypothesis will usually specify the proposed nature of a relationship between two or more **variables**, which are the empirical stand-ins for theoretical concepts. These are the real world **things** that compose our theory.

Variables and their Functional Roles

Theories explain why and how one variable varies depending on the character or value of another variable. This means we have a particular nomenclature to refer to variables given their functional role.

- ▶ 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. The value of the independent variable is expected to be independent of variation in Y.

Operationalization

Adequately building and testing theories requires that we think of real-world ways to measure the variables of interest.

For example, we might be abstractly interested in *socioeconomic inequality*. But, there are **many** ways to measure this concept.

Therefore, when we think of **operationalizing** socioeconomic inequality, we need to choose a fixed, systematic measure of this concept to help us identify instances of socioeconomic inequality.

Theory Construction, An Example

Let's say we have a theory that public approval of politicians is not grounded only in their present political careers. Instead, we conjecture that politician approval is also dependent on their social networks preceding their political careers. Specifically, we expect that politicians with previously public lives receive higher levels of public approval through their political term than those who did not have public lives prior.

- What is a hypothesis that we can derive from this theory?

Theory Construction, An Example

Let's say we have a theory that public approval of politicians is not grounded only in their present political careers. Instead, we conjecture that politician approval is also dependent on their social networks preceding their political careers. Specifically, we expect that politicians with previously public lives receive higher levels of public approval through their political term than those who did not have public lives prior.

- ▶ What is a hypothesis that we can derive from this theory?
- ▶ What variables constitute the hypothesis?

Theory Testing

To conduct hypothesis tests that help put our theory to the test, we need an empirical method to evaluate each hypothesis's veracity with collected 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 numerical analyses to determine whether the accumulated evidence favors the hypothesis proposed or the null. Our default is to *favor the null*, and the burden is on the data to show otherwise.
- ▶ This class will focus specifically on quantitative methods based in probability and statistics.

From Theory to Empirics

- ▶ Each observation or occurrence of our variables of interest is a piece of *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

Sets

A collection of elements.

Types of numerical elements in a set

- ▶ 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 $a + bi$

Elements in a set

You can have a set composed of other *things* too. Such as names of people, places, or other things.

Sometimes you will code these non-numerical values to have numerical representation, other times you will leave the names as is. These distinctions will become more clear moving forward and in your coding instruction.

Measurement

- ▶ Nominal: no mathematical relationship between the variables
- ▶ Ordinal: categorical variable with set, non-numeric relationship between values, order of numbers is meaningful
- ▶ Interval: equidistant from one value to the next, but there is not meaningful zero
- ▶ Ratio (interval-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: Eye color or Name

Measurement: Examples

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

- ▶ Nominal: Eye color or Name
- ▶ Ordinal: Rating degree of agreement with a new public policy on a scale from 0 (disagree completely) to 7 (agree completely)

Measurement: Examples

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

- ▶ Nominal: Eye color or Name
- ▶ Ordinal: Rating degree of agreement with a new public policy on a scale from 0 (disagree completely) to 7 (agree completely)
- ▶ Interval: Most standardized test scores (SAT, ACT, TOEFL, GRE, IQ, etc.)

Measurement: Examples

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

- ▶ Nominal: Eye color or Name
- ▶ Ordinal: Rating degree of agreement with a new public policy on a scale from 0 (disagree completely) to 7 (agree completely)
- ▶ Interval: Most standardized test scores (SAT, ACT, TOEFL, GRE, IQ, etc.)
- ▶ Ratio: Number of tax dollars spent per person in a municipality

Measurement: Examples

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

- ▶ Nominal: Eye color or Name
- ▶ Ordinal: Rating degree of agreement with a new public policy on a scale from 0 (disagree completely) to 7 (agree completely)
- ▶ Interval: Most standardized test scores (SAT, ACT, TOEFL, GRE, IQ, etc.)
- ▶ Ratio: Number of tax dollars spent per person in a municipality
- ▶ What are some other examples of each of these types of measurements of variables?

Sets, Variables, and Sample Spaces

- ▶ How does this make sense in probability and statistics?

Sets, Variables, and Sample Spaces

- ▶ How does this make sense in probability and statistics?
- ▶ We measure and assign values to variables. When we conduct this data accumulation process, how we measure something and the bounds of its measurement are consequential to how we then analyze it.

Sets, Variables, and Sample Spaces

- ▶ How does this make sense in probability and statistics?
- ▶ We measure and assign values to variables. When we conduct this data accumulation process, how we measure something and the bounds of its measurement are consequential to how we then analyze it.
- ▶ Knowing how to characterize variables is important to understand the following concepts:

Sets, Variables, and Sample Spaces

- ▶ How does this make sense in probability and statistics?
- ▶ We measure and assign values to variables. When we conduct this data accumulation process, how we measure something and the bounds of its measurement are consequential to how we then analyze it.
- ▶ Knowing how to characterize variables is important to understand the following concepts:
- ▶ Solution set: set of all solutions to an equation

Sets, Variables, and Sample Spaces

- ▶ How does this make sense in probability and statistics?
- ▶ We measure and assign values to variables. When we conduct this data accumulation process, how we measure something and the bounds of its measurement are consequential to how we then analyze it.
- ▶ Knowing how to characterize variables is important to understand the following concepts:
- ▶ Solution set: set of all solutions to an equation
- ▶ Sample space: set that contains all the values a variable can take

Sets, Variables, and Sample Spaces

- ▶ How does this make sense in probability and statistics?
- ▶ We measure and assign values to variables. When we conduct this data accumulation process, how we measure something and the bounds of its measurement are consequential to how we then analyze it.
- ▶ Knowing how to characterize variables is important to understand the following concepts:
- ▶ Solution set: set of all solutions to an equation
- ▶ Sample space: set that contains 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

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

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: $A = \{x | x \in \mathbb{R}, x < 200\}$

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.”

Characterizing Sets

- ▶ **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 $[\]$.

Characterizing Sets

- ▶ **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 $[\]$.
- ▶ Finite (finite number of elements) or Infinite (no limit)

Characterizing Sets

- ▶ **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 $[]$.
- ▶ Finite (finite number of elements) or Infinite (no limit)
- ▶ Countable or Uncountable: elements can be counted or not (e.g $\{0, 1\}$ is countable)

Characterizing Sets

- ▶ **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 $[\]$.
- ▶ Finite (finite number of elements) or Infinite (no limit)
- ▶ Countable or Uncountable: elements can be counted or not (e.g $\{0, 1\}$ is countable)
- ▶ Ordered and Unordered sets: ordered set has significance in order while in unordered, no significance in ranking

Characteristics of Sets

- ▶ Subset: elements of some sets are contained within a second set, e.g. even integers and integers. (proper subset: \subset , subset: \subseteq)
- ▶ Cardinality: number of elements
 - ▶ Singleton: one element
- ▶ Empty set: no elements.
 - ▶ $A = \emptyset$ is VERY different from $A = \{\emptyset\}$
- ▶ Universal set (the universe: all elements)

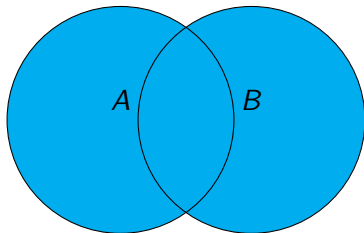
Sets and Set Operations

Now that we know a bit about what sets *are*, we can now visit what to *do* with sets.

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 .

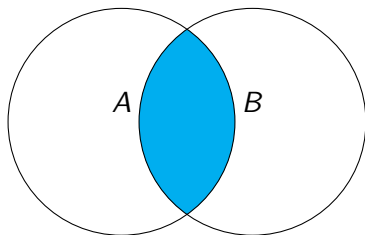


Sets and Set Operations

Intersection: $A \cap B$

Finding the elements in sets or subsets that are shared, i.e. that intersect.

Here we highlight all of the elements in A **AND** B .

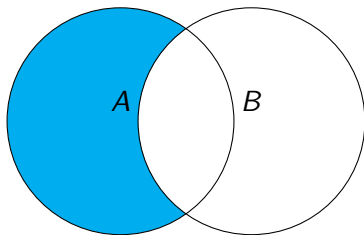


- ▶ Two or more sets are **disjoint** if their intersection is empty, i.e. $A \cap B = \emptyset$. What might the Venn diagram of disjoint sets look like?

Sets and Set Operations

Difference: $A - B$

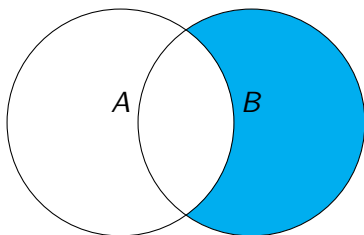
The elements of one set less the elements of another.
Here we show the difference between sets A and B ($A - B$).



Sets and Set Operations

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 .

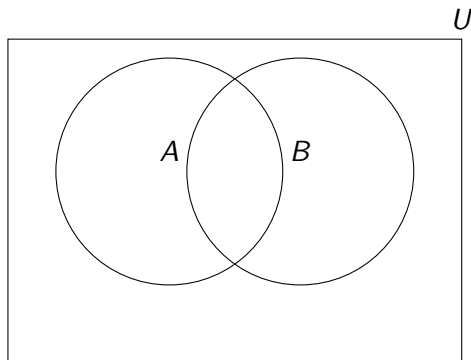


Sets and Set Operations

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 U belong to a subset, either A or B .



Sets and Set Operations

Cartesian Product: $A \times B$

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

A : {trees, apples, rivers}

B : {pencil, paper, chalk}

	trees	apples	rivers
pencil	(pencil, trees)	(pencil, apples)	(pencil, rivers)
paper	(paper, trees)	(paper, apples)	(paper, rivers)
chalk	(chalk, trees)	(chalk, apples)	(chalk, rivers)

Sets: Review

- ▶ Unions: OR, \cup
- ▶ Intersections: AND, \cap
- ▶ Ordered/Unordered
- ▶ Complements (not inside) c , written A^c , for example.
- ▶ Subsets and proper subsets (contained within): \subset, \subseteq
- ▶ Cardinality (number of elements)

Logical and Relational Operators

- Important for notating logical statements in R. Understanding Boolean logic is also important for making library searches!
This will be an easy but effective tool to have at your disposal.

Notation	Meaning
&	AND
	OR
!	NOT
<	LESS THAN
<=	LESS THAN OR EQUAL TO (LEQ)
>	GREATER THAN
>=	GREATER THAN OR EQUAL TO (GEQ)
==	EXACTLY EQUAL TO
!=	NOT EQUAL TO
%in%	IN THE SET

A note on operators

- ▶ The logical AND might also be annotated \times or $*$. The logical OR might also be annotated $+$. The logical operation is still the same.
- ▶ Similar to PEMDAS, there is an order of precedence in these operators. Mathematical operators go first, followed by relational operators (e.g, \leq), then logical operators. You can further segment your syntax with these operators by the use of parentheses, just like in a mathematical expression.