Math Camp Textbook

Jean Clipperton and Sarah Moore

2022-07-27

Contents

Chapter 1

Introduction: Welcome!

This text is a living document that will continue to change and evolve over time. The goals are for this book to provide supplement support and context for the material covered in our course, Math Camp.

The goal is to provide supplemental materials and explanations to accompany the course slides.

1.1 How to read / use this text

Depending on how you prefer to learn, you can read this on its own before or after class sessions (or both!). Please always feel free to check in with any questions you may have, and/or any typos you may find. This helps your learning and improves the text resources!

1.2 Background / getting started:

Over the academic year, you'll encounter empirical questions that authors are working to answer. Our course prepares you to be able to understand how others have answered those questions and to consider how you might go about answering your own questions.

1.2.1 Goals and backgrounds

Many of you may not be sure about whether you'll use quantitative methods and / or of the value they may bring to our research. We want to fully acknowledge that people are coming in with different goals. Our hope is that we can end

Math Camp with people prepared to excel in quantitative coursework regardless of whether their goal is to conduct that research themselves, to be conversant in the literature of their research area, or to explore it as a possible avenue for their future work.

1.3   Additional resources

The materials for this text are based off the book by Moore and Siegel and have been adapted over time with incorporation of additional materials from Jeff Gill. We also try to incorporate and recognize other outside sources as we use them.

Chapter 2

A broad introduction to social science

The goal behind Math Camp is getting you into position to explore topics interesting to you, and to conduct your own research. We recognize that people come from very different backgrounds and have quite different levels of experience, background, and coursework. The goal of this chapter is to provide a basic orientation to why and how we care and think about the pieces that will appear in the rest of this text.

2.1 The scientific method: forever in our hearts

The scientific method is going to underpin a lot of our conversations and will help us formulate our research questions. It is a building block for empirical social science research. There are different ways you can structure / organize it, but here's my preference: + **Puzzle** This is your research question: why / how are we seeing this thing that a) research wouldn't lead us to expect and / or b) how might we understand how two competing pieces / theories can help explain *this* particular phenonmena?

- **Theory** This is our explanation for why / how the pieces come together. It connects the different key concepts (see more on this below) and provides some type of explanatory mechanism (over the term, we'll also get into causal vs associational connections). You will want your theory to be *falsifiable.*

    - *Hypotheses* these are the *testable implications* of your theory.

- **Method / Test** This is how you will go about answering or addressing the research question. Through this method, (e.g. ethnography, regression, etc.), you'll work to discover the answer to your puzzle.

  - Note: there are many, many ways you can choose to answer your research method and while we won't cover this in math camp, this will be part of the discussions in your first year of quantitative methods training.

- **Conclusions** These are the conclusions from evaulating your *theory* with your *method / test* to answer your *puzzle.* You may be able to answer the questions and / or have soem kind of preliminary evidence.

2.2   Puzzles / Research questions

The research question guides your project. This is typically a question that has arisen that has some significance / value in the general realm of your research. NOT ALL QUESTIONS ARE PUZZLES / INTERESTING. Finding and framing your research question can take a frustratingly long amount of time to get 'right.' Some questions are unanswered because the answers are uninteresting while others are unanswered for reasons relating to methods, or because people think they already know the answer. Ideally your research question has some kind of broader interest beyond your particular research area.

2.3   Theories

Theories are how we explain or answer our research question – they provide the why or how. In developing your theory, you may develop, define, or incorporate key ideas or *concepts* (see below for more) that link together in some way.

Typically, your theory is explaining *how* the concepts connect to one another (e.g. do they have a positive relationship? is there some conditional element / connection?). One way to think of your theory as the story or explanation for what is happening, [add example XXX]

2.3.1   Concepts

Concepts are key terms that are meaningfully connected in your theory. These can be specific terms to your research question (e.g. how you're defining 'approval'), or can be more general terms that are coming from the literature on your research topic (e.g. what constitutes an institution).

In any case, these are going to be key ideas that will feature prominently in your research. They are likely part of the research question you're trying to answer.

2.3.2 Operationalization & Measurement: Variables

The concepts you specify will likely be quite large and vague in some way. To turn these concepts into something you can work with in a research project, you will need to *operationalize* them. This means that you'll take a broad or vague concept, such as 'eligible voters' and turn it into something you can measure and work with.

Even in the case of 'eligible voters,' you might mean all people who are over 18 since they should be capable of voting. However, what about registration (especially in states with restrictions around same day registration), criminal records (especially in states with restrictions around incarceration or felony conviction), and past voting history (have they ever voted?). You'll want to be very specific around how you're defining (operationalizing) your key concepts. This process can turn the concept into a variable that you analzye / use in your emipirical research.

2.3.3 Hypotheses

Hypotheses are the testable implications of your theory. Another way to consider this is to ask, 'If my theory were true, what would we expect to see?' These are often more narrow and specific than your theory.

For example, you will likely specify a direction or connection between your concepts (e.g. example XXX).

2.4 Doing research / getting data

Perhaps frustratingly, we're just building up tools and skills here so that in the upcoming year, you'll be ready to apply the concepts you learn to your own chosen topics. Over the course of the academic year, you'll have the opportunity to read academic work and to find and analyze datasets that are interesting to you. Keep these structural elements in mind when you're doing the work to help you decide on a topic and framing for your research question.

Chapter 3

Sets, Notation, and Logic

In this chapter, we'll talk about how to get acclimated to reading quantitatively-focused texts. A lot of the content here won't be directly tested or assessed but is often assumed and 'supporting' material for the content you'll be learning.

Working though this chapter in particular can help orient you to a different way of approaching similar content.

3.1   Sets (preview) and natural numbers

We'll continue to work more with sets (particularly when it comes to probability), but I find it's also incredibly helpful for thinking about things like measurement and operationalization. It may be helpful to picture Venn Diagrams when thinking about sets: you can consider how the different elements come together and connect to one another. Sometimes it's concentric circles (also known as subsets), where each set is completely composed of another.

3.1.1   Measurement and sets

Often, when dealing with a variable, you'll need to determine the level of measurement to use when operationalizing it (see previous chapter for more information on operationalization and variables).

For example, you may be interested in something like ncome, which can be measured yearly, monthly, weekly, or hourly. You can choose to get these amounts in dollar amounts that can be discrete values or to have respondents select bands for their income (e.g. $20,000-$40,000).

3.2   Set Notation

How we discuss and work with sets requires its own notation. Sets are simply collections of elements. There can be conditions that specify how an element comes into a set (see below for solution sets). We can talk about sets in relation to one another, if they have overlapping elements, or no overlap at all. Some people find that the most challenging part of working with sets is the notation.

3.2.1   Symbols: Requirements for being in a set

These symbols talk about what can be in a set, and / or specify the relevant elements.

Notation	Meaning	Use
{	bracket	Specify a set (e.g. $\{2, 3\}$)
$\exists$	"there exists": true for at least one thing	Specify a condition to be satisfied
$\forall$	"for all"; true for all elements	Specify which elements belong in a set (all that satisfy a criteron)
$\exists$	"exists"; there is something true	Specify a rule or proposition that is true
$\in$	"in" or "element of"	States what something / an element is a member of
| Such that	used in set theoretic definitions re: which values satisfy a particular (set of) condition(s)	
$\notin$	excluding (element)	States that something is not a member of a set
$\equiv$	equivalent	set theory equal

For example, $A = \forall x \in$ the set of natural numbers that are divisible by 4. This is telling us that A is a set of numbers that is comprised of natural numbers that are divisible by 4. We could list out the elements of this set. Note that curly brackets are always used for sets, never parentheses. E.g. $\{4, 8, 12, …\}$ and not $(4, 8, 12, …)$.

3.2.2   Symbols: Set operations

Once you have sets, you may want to do certain operations with them, like determine how they relate to other sets. This may sound high-level and abstract

initially, but you can also think of it as wondering how many respondents within a survey satisfy certain criteria and how they relate to members who satisfy other criteria.

Notation	Meaning	Use
$\subset$	Subset	Think of it like the set theory version of less than
$\subseteq$	Subset	Think of it like the set theory version of less than equal too
$\emptyset$	Empty set	Set theory version of zero
$\cap$	Intersection	Set theory version of 'aNd'
$\cup$	Union	Set theory version of or
	Difference	Set theory version of minus to remove elements from sets

3.2.3 Symbols: Sets of numbers

Notation	Meaning	Use
N	Natural numbers	$\{(0), 1, 2, 3, 4, 5, …\}$
Z	Integers (pos and neg)	$\{…, -3, -2, -1, 0, 1, 2, 3, …\}$
Q	Rational numbers (quotient)	(all fractions–produced by numbers divided by an integer)
R	Real numbers (pos, neg, fractions)	(any point on a number line)
I	Imaginary number (i $= \sqrt{(-1)}$)	
C	Complex numbers (a + bi)	

3.2.4 Solution Sets

When considering the collection of possible responses you might collect on a response, you can think of this as a solution set. This might be dollar amounts, other numeric values, or a collection of roles or responses.

More formally, a solution set is a collection of items that satisfies a certain condition. For example, all odd numbers between 2 and 10 would be represented as $\{3, 5, 7, 9\}$.

3.2.4.1 Measurement

How we turn our responses into possible solution sets with different levels of *variable measurement*: this can be at very fine levels, or at larger chunks. For

example, one canonical example is age, which can be a bit tricky. Consider how we might want to measure age: + **Interval-ratio** or **continuous** measurement is measured such that any possible value on the number line is possible. + ex: age could be measured as a fraction, e.g. 246/12, to report a continuous measure + **Discrete**: measurement is whole integers + ex: age could be reported as an integer + **Ordinal**: categorical variable where the the categories have a relationship to one another. Here, it's important that each category be separate from one another and that there is a way to *order* the categories. + ex: age: could be reported as a categorical variable, such as 0-17, 18-34, 35-49, etc. Note that there's a clear way to order the categories and we wouldn't have the options as 18-34, 0-17, and 35-49. That would be the 'wrong' order. + **Nominal**: in this measurement, the variables are categorical, but have no inherent sense of order. For example, eye color options could be arranged any number of ways. + ex: age: here, it's tough to have age as a nominal varible, but we could have it as a binary variable, such as retirement age or not, and that could potentially have any order for the two categories.

3.3   Greek Letters

Greek letters are commonly used in notation in algebra, calculus, and regression. Below, we have a list (via https://www.overleaf.com/learn/latex/List_of_Greek_letters_and_math_symbols) that provides the Greek letter and it's counterpart in the Modern English alphabet.

α A	N
β B	ξ Ξ
γ Γ	o O
δ Δ	π Π
ϵ ε	ρ ϱ P
ζ Z	σ Σ
η H	τ T
θ ϑ Θ	υ Υ
ι I	ϕ φ Φ
κ K	χ X
λ Λ	ψ Ψ
μ M	ω Ω

3.3.1   Commonly used letters

You'll frequently see the following Greek letters in your coursework, so it can be helpful to familiarize yourself with them now:

- δ "delta" (used for integrals and discount factors in game theory)

- Δ "delta" (used for difference/change)
- β "beta" (used for coefficients in regression)
- μ "mu" (used for means)
- σ "sigma" (used for standard deviation)
- λ "lambda" (used for eigenvalues in linear algebra)
- ϵ "epsilon" (used for the error term in regressions)

3.4   Working with variables and letters

Often, we'll have letters to substitute in for our variables, such as x. Typically, 'later' elements in the alphabet (e.g. x, y, z) are used for variables and 'earlier' elements are used for constants (e.g. a, b, c). This is not necessarily a 'rule' per se, but it is a common convention. You'll also see i and n used in sequences and series.

3.5   Logic and Proofs

Here, we review a few relevant terms and symbols.

3.6   Proof terms

You may encounter proofs in texts or materials you read. Here are a few terms that can be helpful to sort out.

- **Assumptions**: statements taken to be true

- **Proposition**: statement thought to be true given the assumptions

- **Theorem**: proven proposition

- **Lemma**: theorem, something of little interest

- **Corollary**: a type of proposition that follows directly from the proof of another proposition and does not require further proof

3.7   Necessary and Sufficient

When exploring outcomes and the connection between conditions, you may wonder how they occur together. One way to explore this is necessary and sufficient conditions. Consider an outcome D and three conditions, A, B, and C.

3.7.1 Sufficient

A sufficient outcome is something that occurs *also* when our outcome variable occurs. So, if both A and B occur and D also occurs, they are *sufficient* for D to occur. Note, this means that D could happen without A or B but we don't see A or B without D.

3.7.2 Necessary

For necessary conditions, we never see the outcome variable without the other effects. So, any time we observe D, C has always also occurred. Note: this means that C could happen without D but never D without C.

3.7.3 Necessary and Sufficient

Necessary and sufficient conditions occur when we never observe the outcome variable without the input variables AND we never observe the input variables without the output. One way to consider things is something is *sufficient* if we sometimes observe our outcome when it happens, but not always. Something is *necessary* if we only see our outcome when it happens, but not always. **Necessary and sufficient** is when we only ever see the outcome when we see the event/condition.

3.8 Logical Operations

We'll also use logical operators to do operations on sets. These can be helpful for data analysis as well, for example, when we talk about filtering observations and tidying our data.

Notation	Meaning	Use
$\wedge$	And	Discussing elements in both sets
$\vee$	Or	Discussing elements that are in multiple sets
$\sim$	Not	Negates the equation
!	Not	Negates the equation (used for R most often)
$<$	Less than	Inequality(good for specifying conditions when filtering)

Notation	Meaning	Use
<=	Less than equal to	Inequality (good for specifying conditions when filtering; include the value)
>	Greater than	Inequality (good for specifying conditions when filtering)
>=	Greater than equal to	Inequality (good for specifying conditions when filtering; include the value)
!=	Not equal to	Exclude values when filtering (anything other than the exact value)
==	Exactly equals	Helpful in R for when a value is exactly satisfied
%in%	In	Useful when searching for terms

3.9 Proofs

In the texts you read this year, you may encounter proofs of work: direct and indirect. We won't do proofs in Math Camp, but we'll reference them implicitly throughout the course and in your fall coursework.

3.9.1 Direct proofs

Direct proofs work through proof and typically use one of the following methods: + **General (deductive) proof**: typically done using definitions, etc. Showing how the outcome logically follows building on rules and assumptions. + **Proof by exhaustion**: Break up the outcome into sub cases and show for each case (done often in game theory for possible values) + **Proof by construction**: These proofs demonstrate existence (is there a square that is the sum of two squares?). + **Proof by induction**: Start small and show it is true for any number (e.g. start with a small n, n=1, then expand to n+1)

3.9.2   Indirect proofs

These show that the outcome must be true because there is no possible alternative.These are typically demonstrated using the following methods:

- Proof by counterexample: using a counterexample (x implies y, yet we observe y without x...x cannot imply y (aka x not necessary for y)).
- Proof by contradiction: assume that the statement is false and try to prove it wrong, eventually demonstrating that a contradiction emerges. Thus, the statement cannot be false.

3.10   Conclusion

This chapter covered a variety of topics around sets, notation, and proofs. We worked through key symbols and terminology that will be useful as you navigate through Math Camp and into the first year of quantitative methods coursework. It may be helpful to come back and revisit these concepts over time – we've tried to be clear about when and where you'll use them and how they will come to play into future concepts.

Chapter 4

Sequences and Sets

Sets allow us to talk more formally about collections of items. We can construct sets formally, as we'll see in this chapter.

4.1 Sequences

A sequence is an ordered list of numbers. They can be infinite or finite but all are *countable*. When discussing items in the sequence, we can refer to them by their position, for example whether they are the first or third element (e.g., for the sequence $\{1, 3, 5, 7, 9\}$, x_2 is 3). One thing to note is that we typically number the elements starting from 1 (e.g. x_1, x_2, ..., etc.), but in some programming languages, like Python, the first element is referred to as 0. This is just something to keep in mind in the future.

You can generate sequences using an equation or formula, for example, if you have an integer that ranges from 1 to 5 and your sequence follows the pattern of 2x+3, you could write it as $\{2x + 3\}_{x=1}^{5} = \{5, 7, 9, 11, 13\}$. The fourth element of this would be when $x = 4$, or x_4, and is 11. (note that you can use any variable for the sequence. I've used x here, but you could use a different letter).

4.2 Series & Summation

A series is the *sum* of a sequence. Working with summation symbols can offer a way to avoid tedious by-hand calculations.

For example, if we wanted to add up the numbers from 1 to 8, we could write out, $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8$. When writing this out and adding them together, we'd find that it equals 36. Alternatively, we could write this out as

a sum and use a simple equation to find the solution. While this may not be much of an effort saver for our example, imagine you were adding up 50 or 100 numbers and the value becomes clearer.

Summation takes the form of having our summation operator, Σ, for some value or function. For example, when adding from 1 to 8, we would represent this as $\sum_{n=1}^{8} n$. It can be represented inline like you see here, or with the limits, above/below, like this example:

$$\sum_{n=1}^{8} n \tag{4.1}$$

In either case, the general format is to have the starting value below the summation symbol Σ, the ending value above the symbol, and the equation for generating the series to the right of the symbol.

4.3 Sets

As we've discussed, sets provide an opportunity to have a collection of values / objects. We can then perform operations on these sets. We can refer them through **roster notation** where we simply list the elements or through **set builder notation** where we provide a kind of 'recipe' for how to generate the elements. For example, the **set builder** notation for a set could be something like, $\{x | 0 < x \leq 10, x \neq 5\}$, or we could provide it in roster notation, $\{1, 2, 3, 4, 6, 7, 8, 9, 10\}$.

4.3.1 Details, sets

There are two types of sets: **finite** (have a finite number of elements) and **infinite** (no limit on the number of elements). They can be **countable** (elements can be counted) or **uncountable** (elements cannot be counted). Not only do we care about sets and how elements are contained within them, we care about the boundary of the sets. **Open sets** are sort of like the set theory version of open brackets (). In contrast, **closed sets** have a clear boundary, like with square brackets, [].

We care about the number of elements (cardinality) of a set (a **singleton** set means there is only one element), while an **empty set** has no elements and is represented like this $\emptyset$.

Sets can be **ordered** (like a preference ranking) or **unordered** (meaning that the oardering of elements has no external or substantive meaning). When talking about all possibilities, or a set that contains everything, we talk about a **universal set**.

4.3.2   Relating sets to each other

We can start thinking about how sets relate to one another, specifically how the composition of elements within a set compare to one another. To help work through our definitions, we're going to start with five example sets:

- Set A: $\{1, 2, 3, 4\}$
- Set B: $\{1, 2, 3, 4\}$
- Set C: $\{1, 2, 3, 4, 5, 6\}$
- Set D: $\{6, 7, 8\}$
- Set E: $\{10, 20, 30, 40\}$

4.3.2.1   Union, Disjoint, Complements, Partitions

We can think about bringing sets together, what they share in common, and what they don't. The first way to think about comparing sets is the **union** of two sets, $\cup$. One way to think about this is that it's like 'or' in that anything that is in EITHER of your sets will be in the union of the two sets. You can have the union of multiple sets. For example, $A \cup E = \{1, 2, 3, 4, 10, 20, 30, 40\}$ and $A \cup C \cup E = \{1, 2, 3, 4, 5, 6, 10, 20, 30, 40\}$. Note that we don't duplicate elements when they're in both sets.

We might also be curious what is in BOTH sets, meaning the shared elements. This is the **intersection** of sets and can also be considered across multiple sets. For example, $A \cap B = \{1, 2, 3, 4\}$ but $A \cap E = \emptyset$ (note that because the intersection between A and E is empty, we don't have $\{\emptyset\}$ but instead $\emptyset$ because the $\emptyset$ is like the set theory version of zero). In the circumstance where the intersection is empty, we say that those sets are **disjoint**. In our example, sets A and E are disjoint sets. We can also say that these sets are **mutually exclusive**.

We can consider the **difference** of sets where we have what is effectively the *remainder* from one set when looking at the intersections of sets. One way to think about this is that it's going to be . $x \in A \wedge x \notin B$ For **disjoint** or **mutually exclusive sets**, the difference will be the original set. So, $A \ E = A$ but $C \ A = \{5, 6\}$. Approaching this on a larger scale, if we're interested in everything not in any one set or combination of sets, that is called the **complement**. This is a set that contains all possible values that are not in the original set / item.

We talked about the universal set. For our set examples above, we haven't discussed all possible values that could be an element of a set. If, for example, our universal set is all numbers from 1 to 40, then the complement of A would be $\{5, 6, 7, \ldots, 40\}$.

In considering sets, we could want to split up or divide the elements. This is called **partitioning a set**. When we do this, we divide a set into smaller disjoint sets. This can be valuable if you're applying treatment to groups in an

experiment or thinking about how groups may differ from one another and the individuals can be in only one group.

4.3.2.2 Subsets

Starting from these examples, we can now begin with the idea of subsets. A **proper subset** is when one set is contained within another (think: set theory version of less than ($<$)). It is represented like this $\subset$. In our above example, set A and set B are each a proper subset of set C ($A \subset C$ and $B \subset C$). However, A and B are not proper subsets of each other because the share the exact same elements. In contrast, A and B **are** *subsets* of one another (just not proper subsets). Note that C, D, and E are not subsets of one another.

4.3.2.3 Cartesian Products

At times, you may want to combine elements of two sets, possibly for pairs or for thinking about treatment groups. This is a **Cartesian product** whereby we take every element from set A and match it in a pair with every element from set B.

4.4   Conclusion

In this chapter, we've talked about series, sequences, summation, and sets. We've covered how to think about generating sets, how to relate sets to one another, and possible options for understanding how sets come together.