PHS SummR Camp: Introduction to Mathematical Notation

Jarvis T. Chen (jarvis@hsph.harvard.edu)

August 25, 2020

PhD in Population Health Sciences
Harvard T. H. Chan School of Public Health

- **Mathematical notation** uses symbols to represent mathematical objects and ideas.
- Becoming familiar with mathematical notation helps you to read and understand the methodological literature and to communicate efficiently and precisely.

For population health scientists working with quantitative data, when mathematical notation works well:

- it helps us to specify in precise terms how we use the data at hand to learn about population quantities and relationships
- it explicitly encodes the assumptions we need to draw inferences
- it illuminates relationships between the different kinds of quantities we observe
- it makes it possible to communicate our ideas and methods to our colleagues in a transparent way
- it makes it possible for others to replicate our work

When mathematical notation does not work well:

- it obscures relationships and creates confusion
- it assumes conventions and usage that are not explicitly stated, and therefore not universally intelligible
- it makes it difficult for others to replicate our work
- it makes us frustrated and makes us question our understanding

Unfortunately, there is no such thing as a “perfect” notational system.

- notational conventions differ across disciplines
- often we find ourselves having to amend our notation in order to highlight particular aspects of the specific problem we are working on
- it is always important to be able to translate back and forth between our conceptual understanding in words and our formal notation in mathematical symbols

In this week's PHS SummR Camp, we'll be reviewing some of the key concepts you'll need for PHS2000. In the course of this, you'll see many examples of mathematical notation, including many of the conventions that are used in biostatistics, epidemiology, econometrics, and the quantitative social sciences.

When you see notation that is confusing to you, please ask us about it!

Vectors, Scalars, & Matrices

Objectives

- See how vectors, scalars, and matrices can be used to refer to sets of numerical values
- Learn the notational conventions for vectors, scalars, and matrices.

Conventions: Vectors, Scalars, & Matrices

As quantitative population health scientists, we find ourselves spending a lot of time thinking about

- numbers (the “quantitative“ part)
- not just single numbers but sets of numbers (the “population” part)
- different sets of numbers referring to different variables of interest

Conventions: Vectors, Scalars, & Matrices

We need an efficient way of referring to these sets of numbers, e.g.

- all of our study participants' ages, instead of just one subject's age
- all of the variables measured on a single subject (e.g. age, race/ethnicity, income, cholesterol, body mass index)
- all of the variables measured on all of the subjects in our study

This becomes very important when thinking about how we can manipulate these sets of numbers to learn something about the population from which they come.

Conventions: Vectors, Scalars, & Matrices

- A **vector** is a structured set of inputs (e.g., numbers), arranged in a list.
  - e.g. $(1, 2, 4, 3, 5)$ (row vector)
  - e.g. $\begin{pmatrix} 2 \\ 5 \\ 13 \\ 9 \\ 4 \end{pmatrix}$ (column vector)
- By convention, vectors are often represented by lower case Roman letters in **boldface**, e.g.

$$\mathbf{x} = (1, 2, 4, 3, 5)$$

- A **scalar** is a vector with just one element, usually represented by a non-boldface lower case Roman letter, e.g. $x = 7$.

Conventions: Vectors, Scalars, & Matrices

- A **matrix** is a structured set of *vectors* all with the same length (number of elements)
- e.g.

$$\mathbf{X} = \begin{bmatrix} 5 & 3 & 6 & 9 \\ 4 & 12 & 3 & 13 \\ 8 & 2 & 0 & 19 \end{bmatrix}$$

- A matrix is usually represented by an upper case Roman letter in **boldface**.

See End-of-Summer Camp: Math Review for more details on vectors and matrices!

Conventions: Vectors, Scalars, & Matrices

Note: Sometimes when working with a lot of vectors and matrices, some authors will suppress the boldface notation for simplicity. Usually this will be obvious from how vectors and matrices are defined in their notation.

Indexing

Objectives

- See how indexing can help us refer to specific elements within a vector or a matrix.
- Appreciate how multiple subscripts can refer to elements of a matrix.
- Note the conventions for referring to the number of subjects in a dataset (n) and the number of variables in a regression model (p).
- See how indexing works as part of the summation and product operators.

Conventions: Indexing

When working with population data, we will often observe multiple values of the same variable, e.g. over subjects, over time, etc. We will often use indexing notation to represent multiple values of the same variable, e.g.

- X is height in cm, which we observe in a sample of 300 women. We can represent the value for person i as x_i, for $i = 1, \ldots, 300$.
- All of the observed x's in the study form a vector $\mathbf{x} = (x_1, \ldots, x_{300})$.
- x_{241} refers to the 241st element in the vector $\mathbf{x}$.
- x_i refers generically to the ith element of $\mathbf{x}$.

Conventions: Indexing

- We ask 100 subjects in our study to record the number of hours they sleep each night for a month. Thus, each of the 100 subjects in our study has 31 records of the number of hours they slept each night, which we could index as y_{it} where i indexes subject ($i = 1, \ldots, 100$) and t indexes night ($t = 1, \ldots, 31$).
- The observations for each individual i are a vector of length 31, $\mathbf{y}_i = (y_{i1}, \ldots, y_{i,31})$
- All of the observations in the study form a matrix Y of dimension 100×31,

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1,31} \\ y_{21} & y_{22} & \cdots & y_{2,31} \\ \vdots & \vdots & \ddots & \vdots \\ y_{100,1} & y_{100,2} & \cdots & y_{100,31} \end{bmatrix}$$

Conventions: Indexing

When working with a dataset,

- often we will use n to represent the number of subjects in our sample, so we might use i to index subjects with $i = 1, \ldots, n$.
- often we will use p to represent the number of variables included in a model, e.g. the model includes variables $X_1, X_2, \ldots, X_p$.

Obs	X_1	X_2	...	X_p
1	x_{11}	x_{12}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	x_{n1}	x_{n2}	...	x_{np}

Conventions: Summation & Products

We use indexing when using the summation operator:

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

means sum all of the values of x_i from $i = 1$ to $i = n$.

We can also sum over multiple dimensions, e.g.

$$\sum_{i=1}^{n} \sum_{j=1}^{p} x_{ij}$$

means sum over all the values of x_{ij} (from $i = 1$ to $i = n$ and $j = 1$ to $j = p$).

Conventions: Summation & Products

Similarly,

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

means take the product of all of the values of x_i from $i = 1$ to $i = n$.

Conventions: Summation & Products

Sometimes, when there are multiple dimensions, to keep the number of letters being used under control, people will represent the maximum index with a capital letter, e.g.

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk}$$

for $i = 1, \ldots, I$, $j = 1, \ldots J$, and $k = 1, \ldots, K$.

Note that in this case, the capital letter is **not** denoting a random variable! See Slide 27 for that convention.

Conventions: Summation & Products

As an alternative to specifying summation from a starting point to a stopping point is to use set notation, e.g.

$$\sum_{i \in S} x_i$$

means sum all values of x_i where i is in the set S.

Conventions: Summation & Products

Sometimes, you will see the notation $\sum_i x_i$ when it's clear that summation is over all possible values of i.

Note that this enables us to write the double summation more compactly as

$$\sum_{i,\,j} x_{ij}$$

Another example (from probability, the definition of expected value):

$$\mathbb{E}[X] = \sum_{x} x\, f(x)$$

Here the subscript x on the summation symbol means "over the range of x."

See PHS SummR Camp: Probability for more on the definition of expected value!

Poll 1

Set Notation

Objectives

- Learn about special symbols for important sets of numbers.
- See how we can enumerate the elements of a set explicitly using $\{\}$.
- See how the ellipsis $(\ldots)$ can be used to denote elements in a regular sequence.
- Review the meaning of common operators in set notation.

Conventions: Set Notation

A **set** is a collection of elements or members. Typically, we represent a set with a capital letter, e.g. A, B, C, etc. By convention, particular symbols are reserved for the most important sets of numbers:

- $\emptyset$ – empty set
- **N** or $\mathbb{N}$ – natural numbers (non-negative integers)
- **Z** or $\mathbb{Z}$ – integers
- **Q** or $\mathbb{Q}$ – rational numbers
- **R** or $\mathbb{R}$ – real numbers

Conventions: Set Notation

A set can be described by enumerating all of its elements between curly brackets, e.g.

- $\{7, 3, 15, 31\}$ is a set holding the four numbers 3, 7, 15, and 31.
- $\{a, c, b\}$ is the set containing 'a', 'b', and 'c'.

When denoting a set that contains elements from a regular sequence, we sometimes use an ellipsis, e.g.

- $\{1, 2, 3, \ldots, 100\}$ is the set of integers between 1 and 100 inclusive.

Conventions: Set Notation

A few important symbols:

Symbol	Symbol Name	Meaning
$A \cap B$	intersection	objects that belong to set A **and** set B
$A \cup B$	union	objects that belong to set A **or** set B
$A \subseteq B$	subset	A is a subset of B. Set A is included in set B.
$A \subset B$	strict subset	A is a subset of B, but A is not equal to B
$A \nsubseteq B$	not subset	set A is not a subset of set B
$a \in A$	element of	set membership
$x \notin A$	not element of	no set membership

Probability

Objectives

- Learn basic probability notation
- Learn the conventions for random variables and their realizations.
- Note the notation for common operators for random variables (expectation, variance, etc.)

Conventions: Probability

- $P(A)$ is the probability of event A occurring. Sometimes written as $\mathbb{P}(A)$ or $Pr(A)$.
- $P(A \cap B)$ is the probability that events A and B both occur.
- $P(A \cup B)$ is the probability of either event A **or** event B occurring (where "or" means one or the other or both)
- $P(A|B)$ is the conditional probability of event A occurring **given** that B has occurred.

Conventions: Probability

- **Random variables** are usually written with upper case roman letters: X, Y, etc.
- Particular **realizations** of a random variable are written in corresponding lower case letters. e.g. $P(X = x)$ is the probability that the random variable X is equal to the specific value x.
- We will often assume that a random variable X follows a particular **probability distribution**, e.g.
  - $X \sim Normal(\mu, \sigma^2)$
  - $X \sim Bernoulli(\pi)$
  - $X \sim Poisson(\lambda)$
- Note that each of these distributions has one or more **parameters** associated with it.

Conventions: Probability

- Probability density functions (pdf) and probability mass functions (pmf) are denoted by lowercase letters, e.g. $f(x)$ or $f_X(x)$
- Cumulative distribution functions (cdf) are denoted by uppercase letters, e.g. $F(x)$ or $F_X(x)$.
- The **joint probability distribution** of random variables X and Y is denoted as $P(X, Y)$, while the joint probability mass function or probability density function is $f(x, y)$ and joint cumulative distribution function is $F(x, y)$

More about pdf's, pmf's, and cdf's on Wednesday!

Conventions: Probability

- Some common operators:
  - $E(X)$ or $\mathbb{E}(X)$ is the **expected value** of X.
  - $\text{Var}(X)$ is the **variance** of X.
  - $sd(X)$ is the **standard deviation** of X.
  - $\text{Cov}(X, Y)$ is the **covariance** of X and Y.
- X is independent of Y is often written $X \perp\!\!\!\perp Y$ or $X \perp Y$.

Poll 2

Population Parameters & Sample Statistics

Objectives

- Note the conventions for distinguishing between population parameters and sample statistics.
- Learn about the distinctions between estimand, estimator, and estimate.
- Appreciate the 'hat' notation for distinguishing between estimand (population parameter) vs. estimator or estimate.

Conventions: Population Parameters & Sample Statistics

Often we will draw a distinction in notation between **population parameters** and **sample statistics**.

Population parameters are "true", usually unobserved characteristics of the population, which we represent with Greek letters, e.g.

μ_X population mean of X

σ_X^2 population variance of X

ρ_{XY} population correlation of X and Y

Conventions: Population Parameters & Sample Statistics

We distinguish them from **sample statistics**, which are estimates of population parameters we compute from the data we have observed in a sample:

Parameter	Sample Statistic
μ_X	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
σ_X^2	$s_X^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
ρ_{XY}	$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$

Conventions: Estimand vs. estimator vs. estimate

This distinction between unobserved population parameters and observed sample statistics (functions of our data) is one we will come back to again and again in PHS2000.

Some other vocabulary worth knowing:

- The **estimand** is the population quantity of interest whose true value you want to know.
- An **estimator** is a method for estimating the estimand.
- An **estimate** is a numerical estimate of the estimand that results from the use of a particular estimator.

Estimand vs. estimator vs. estimate: an example

We run a randomized clinical trial of a new antidepressant and measure each participant's CESD depression scale score at 12 weeks.

- Estimand: θ the true population difference in CESD score under treatment and placebo
- Estimator: $\hat{\theta} = \overline{\text{CESD}}_{\text{control}} - \overline{\text{CESD}}_{\text{treated}}$
- Estimate: $\hat{\theta} = 1.2$, i.e. the actual number we calculate from our data using the estimator specified above.

Conventions: Estimand vs. estimator vs. estimate

Note a few things here:

- Our use of the Greek letter θ to represent the **estimand**, i.e. a population parameter
- The 'hat' notation $\hat{\theta}$ used to represent both the **estimator** (i.e. the numeric recipe for estimating the estimand using the data at hand) and the **estimate** (i.e. the actual value we obtain by running this numeric recipe).
- By convention, θ is often used in statistics to represent a generic parameter (could be a mean, a variance, a regression coefficient, etc.).

Notation in regression models

Objectives

- Remind ourselves of the conventional (generic) notation for linear regression model
- Appreciate how we might amend notation to highlight particular aspects of a model.
- Appreciate why bucking convention in notation can become confusing.
- For reference: common Greek symbols
- A few notes on potentially confusing situations
- Note the trade-offs between using letters vs. names for variables.

Conventions: Regression models

You are probably familiar with seeing β's in a regression model, e.g.

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon_i$$

These β's are unknown population parameters which we will estimate by fitting a model, yielding a set of estimates $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$.

Conventions: Regression models

Imagine that we are writing the methods section of a paper exploring social and environmental predictors of low birthweight. In a sample of births, we have collected data on

- low birthweight (Y_i)
- 4 variables on maternal socioeconomic position (e.g. X_1=education, X_2=household income, X_3=wealth, and X_4=neighborhood poverty)
- 3 environmental variables (Z_1=PM2.5, Z_2=black carbon, and Z_3 =water quality)
- 3 demographic variables (W_1=maternal age, W_2=marital status, W_3=maternal race/ethnicity)

After extensive analysis, we decide to present the results of a regression model that includes all 10 of these variables.

Conventions: Regression models

We could write our model generically, using $p = 4 + 3 + 3 = 10$ to index all of the variables we included:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon_i$$

Or, for more clarity and to draw the reader's attention to the different sets of socioeconomic, environmental, and demographic variables we are including in the model, we could write

$$Y_i = \alpha + \sum_{j=1}^{4} \beta_j x_{ij} + \sum_{k=1}^{3} \gamma_k z_{ik} + \sum_{l=1}^{3} \lambda_l w_{il} + \epsilon_i$$

Notice how we now have different Greek letters representing different parameters of the model: α is the overall intercept, β's represent the effects of the socioeconomic variables, γ's represent the effects of the environmental variables, and λ's represent the effects of the demographic variables.

Conventions: Regression models

Consider the pros and cons of how we amended the conventional notation here:

Pros:

- Our notation highlights the conceptual distinctions between socioeconomic, environmental, and demographic variables by using different letters to represent the variables and different Greek letters to represent the regression parameters.
- The notation is actually more specific than the $\beta_0, \ldots, \beta_p$ notation (which is highly generic).

Cons:

- The reader has to mentally orient to three different variable types (X, Z, and W), three different indexes (j, k, and l), and five different Greek letters (α, β, γ, λ, and ϵ)!

Conventions: Regression models

Consider what we **didn't** do here, e.g.

$$N_z = \sigma_0 + \sigma_1 d_1 + \sigma_2 d_2 + \cdots + \sigma_r d_q + \zeta_z$$

where N_z is the birthweight for subject $z = 1, \ldots, Z$, $d_1, \ldots, d_q$ are the covariates, $\sigma_0, \ldots, \sigma_q$ are the regression coefficients, and ζ_z are the error terms.

There is nothing stopping us from defining and using this notation to describe our model, but the notation here is so removed from convention for regression models that it would be virtually unintelligible to most readers!

Conventions: Regression models

In general:

- Y used to represent the outcome
- X used to represent predictors
- Sometimes X used to represent exposures of particular interest and Z used to represent covariates ('control' variables)
- In time to event data (survival analysis), sometimes T is used to represent event times
- Usually β used to represent regression coefficients (occasionally we will also see α, γ, δ, λ, τ...)

For reference: lower case Greek letters

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

For reference: upper case Greek letters

(In practice, Greek letters that have Latin look-alikes are not used to avoid confusion).

Γ	Gamma
Δ	Delta
Λ	Lambda
Φ	Phi
Π	Pi
Ψ	Psi
Σ	Sigma
Θ	Theta
ϒ	Upsilon
Ξ	Xi
Ω	Omega

A few notes

Things to watch out for:

- π is often used in biostatistics to represent a probability, (e.g. $Y \sim Bernoulli(\pi)$ or in a logistic regression model, $\text{logit}(\pi) = \beta_0 + \beta_1 x_1$)
- **BUT** in the normal probability density function,

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

Here, π is the mathematical constant, $\pi \approx 3.14159$!

A few notes

- We saw before that $\sum$ is used as the summation operator, e.g. $\sum_{i=1}^{n} x_i$, and it looks like a capital Σ. But note that Σ is sometimes used to represent a variance-covariance matrix, e.g.

$$\mathbf{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- Don't worry about understanding the details of these usages right now; the message is that sometimes the same symbol is used to represent different things, depending on the context. We'll try to point out when this happens, in order to head off any confusion!

Conventions: Regression Models

Sometimes, to improve readability and interpretability, some researchers prefer to use the names of variables in writing out regression models, instead of representing with Y's and x's. E.g.

$$\mathbb{E}(\text{Birthweight}_i) = \beta_0 + \beta_1 \text{maternal education}_i + \beta_2 \text{age}_i$$

This is not a bad idea, since it helps the reader to connect variables to the model more readily. However, be careful about using abbreviations for variable names (e.g. "edu" for "maternal education") if the abbreviations are not readily interpretable. Long variable names can also become unwieldy!

This convention works well for describing a particular model using particular variables; it is less relevant for describing a methodology, when greater generality is desired.

Poll 3

Notation survival tips

Notation survival tips

- We realize that learning new notation can feel daunting.
- We hope that you'll come to see that there is elegance and even beauty in being able to represent quantitative concepts in rigorous mathematical symbology
- Over the course of this week, we'll be reviewing many of the most common mathematical concepts needed for our course, along with their notation.

Notation survival tips

- When learning new notation, try to take notes on all the different parts of the notation so that you understand what each symbol means.
- When in doubt, look for the part of the lecture slides (or the textbook or the journal article) where the notation is defined. (Most good scientific writing will include this *somewhere*, even if it's just in an appendix).
- Familiarize yourself with the most common conventions used in your field(s).
- When in doubt, ASK!

Writing math notation

Writing math notation

In this course, you will often have to write out math when completing problem sets or take home exams or even when taking notes. What are some options for streamlining your workflow in writing math?

Writing math notation

- You always have the option of writing out math by hand, taking a photo with your phone, and submitting the photo along with your problem set or take home exam.
  - **Pros**: easy to do
  - **Cons**: tedious to have to take a photo every time you need to show math. Image files can be large. Doesn't look elegant.

Writing math notation

- Another option is to use Microsoft Equation Editor in MS Word.
  - **Pros**: most people are already familiar with MS Word and Equation Editor.
  - **Cons**: very "fiddly"; takes a long time to typeset math in a WYSIWYG environment.

Writing math notation

- Another option is to use LaTeX.
  - LaTeX(pronounced *LAH-* or *LAY-tek*) is a syntax language that is used to typeset documents. It's great for creating reports and presentations that look professional and is particularly useful when typesetting mathematical expressions. LaTeX is also great for reproducibility, since there is a "script" file that documents how the output file (generally a .pdf) was created and styled.
  - LaTeXis WYSIWYM (what you see is what you mean)
  - **Pros**: very powerful: you can typeset just about anything. Looks super elegant. LaTeX is the preferred method of typesetting used in any math-related field (e.g. biostatistics), so if you will be working in a related space it's probably worth it to learn.
  - **Cons**: the learning curve can be a bit steep at first. Proficiency in LaTeX is a lifelong journey

Options for using LaTeX

Overleaf

- Overleaf is an online LaTeX edition tool that allows you to create LaTeX documents directly in your web browser.
- Harvard University is providing free Overleaf Professional accounts to all students, faculty, and staff. Overleaf Professional accounts provide real-time track changes, unlimited collaborators, and full document history. You can claim your Overleaf Professional account at `https://overleaf.com/edu/harvard`
- Harvard Library offers a guide to help you use Overleaf at `https://guides.library.harvard.edu/overleaf`
- Also check out this guide by Overleaf.

Homework

We'd like you to set up your Overleaf Professional account before tomorrow's session, as we'll be using it a few times this week. If you haven't used Overleaf or LaTeX before, don't worry: you just need to set up the account and we will give you further instructions tomorrow.

Of course, if you'd like to play around with before tomorrow's session, you are also welcome to do that!

Options for using LaTeX

R Markdown

- R Markdown provides an authoring framework for data science. You can use a single R Markdown file to both
  - save and execute code, and
  - generate high quality reports that can be shared with an audience.
- R Markdown enhances reproducibility since both the computing code and narratives are in the same document, and results are automatically generated from the source code.
- It is worth learning how to use R Markdown to streamline your workflow.
- The best reference source is: `https://bookdown.org/yihui/rmarkdown/`

Options for using LaTeX

- R Markdown provides an authoring framework for data science. You can use a single R Markdown file to both
  - save and execute code, and
  - generate high quality reports that can be shared with an audience.
- R Markdown enhances reproducibility since both the computing code and narratives are in the same document, and results are automatically generated from the source code.
- It is worth learning how to use R Markdown to streamline your workflow.
- The best reference source is: `https://bookdown.org/yihui/rmarkdown/`
- Some additional resources are available here.

R Markdown

- To use R Markdown, you should have installed R (https://www.r-project.org) and the RStudio IDE (https://www.rstudio.com). Next, you can install the `rmarkdown` package in R:

```
# Install from CRAN
install.packages('rmarkdown')
```

- If you want to generate PDF output, you will need to install LaTeX. For R Markdown users who have not installed LaTeX before, you can install TinyTeX (https://yihui.name/tinytex/). TinyTeX is a lightweight, portable, cross-platform, and easy-to-maintain LaTeX distribution.

```
install.packages('tinytex')
tinytex::install_tinytex()  # install TinyTeX
```

- With the `rmarkdown` package, RStudio/Pandoc, and LaTeX you should be able to compile most R Markdown documents.

Options for using LaTeX

R Sweave

- Sweave is another way of integrating R code and output with LaTeX typesetting.
- Like R Markdown, Sweave supports a reproducible workflow by integrating data analysis code with narratives.
- Matt has written a helpful **Guide to LaTeX and Sweave** which we've posted on the SummR Camp tab on Canvas.

Takehome Message

- It's worth taking the time to learn how to use to typeset math, particularly if you are going to use tools like R Markdown or Sweave to create a reproducible workflow.
- However, the learning curve can be steep.
- If your code doesn't work, you can often Google a solution, but also consider posting questions to the Canvas discussion board.
- Take your time learning to use these new tools – remember that there are always other options.
- **Ask for help when you need it.**

Poll 4