Methods Camp

UT Austin, Department of Government

Andrés Cruz and Matt Martin

2023-08-10

Table of contents

Class schedule

Date	Time	Location
Thurs, Aug. 10	9:00 AM - 4:00 PM	RLP 1.302E
Fri, Aug. 11	9:00 AM - 4:00 PM	RLP 1.302E
Sat, Aug. 12	No class	-
Sun, Aug. 13	No class	-
Mon, Aug. 14	9:00 AM - 4:00 PM	RLP 1.302E
Tues, Aug. 15	9:00 AM - 4:00 PM	RLP 1.302E
Weds, Aug. 16	9:00 AM - 4:00 PM	RLP 1.302E

On class days, we will have a lunch break from 12:00-1:00 PM. We'll also take short breaks periodically during the morning and afternoon sessions as needed.

Description

Welcome to Introduction to Methods for Political Science, aka "Methods Camp"! In the past our incoming students have told us their math skills are rusty and they would like to be better prepared for UT's methods courses. Methods Camp is designed to give everyone a chance to brush up on some skills in preparation for the Stats I and Formal Theory I courses. The other goal of Methods Camp is to allow you to get to know your cohort. We hope that struggling with matrix algebra and the dreaded chain rule will still prove to be a good bonding exercise.

As you can see from the above schedule, we'll be meeting on Thursday, August 10th and Friday, August 11th as well as from Monday, August 14th through Wednesday, August 16th. Classes at UT begin the start of the following week on Monday, August 22nd. Below is a tentaive schedule outlining what will be covered in the class, although we may rearrange things a bit if we find we're going too slowly or too quickly through any of the material.

Course outline

Friday morning: R and RStudio

- Introductions
- RStudio (materials are on the website as zipped RStudio projects)
- Objects (vectors, matrices, data frames)
- Basic functions (mean(), length(), etc.)

02 Friday afternoon: Tidyverse basics I

- Packages: installation and loading (including the tidyverse)
- Data wrangling with dplyr (basic verbs, including the new .by = syntax)
- Data visualization basics with ggplot2
- Data loading (.csv, .rds, .dta/.sav, .xlsx)
- Quarto fundamentals

03 Monday morning: Matrices

- Matrices
- Systems of linear equations
- Matrix operations (multiplication, transpose, inverse, determinant).
- Solving systems of linear equations in matrix form (and why that's cool)
- Introduction to OLS

04 Monday afternoon: Tidyverse basics II

- Data merging and pivoting (join(), pivot())
- Value recoding (if_else(), case_when())
- Missing values
- Data visualization extensions: facets, text annotations

05 Tuesday morning: Functions and loops

- Functions
- For-loops and lapply()
- Finding R help (help files, effective Googling, ChatGPT)

06 Tuesday afternoon: Calculus

- Limits (not sure how to teach this in an R-centric way yet, but there must be a way)
- Derivatives (symbolic, numerical, automatic)
- Integrals

07 Wednesday morning: Probability

- Concepts: probability, random variables, etc.
- PMF, PDF, CDF, etc.
- Distributions (binomial, normal; different functions in R and how to use them)
- Expectation and variance

08 Wednesday afternoon: Simulations

- Simulations (ideas, seed setting, etc.)
- Sampling
- Bootstrapping

09 Thursday morning: Text analysis

- String manipulation with stringr
- Simple text analysis (counts, tf-idf, etc.) with tidytext and visualization

10 Thursday afternoon: Wrap-up

- Project management fundamentals (RStudio projects, keeping raw data, etc.)
- Self-study resources and materials
- Other software (Overleaf, Zotero, etc.)
- Methods at UT

Contact info

If you have any questions during or outside of methods camp, you can contact Andrés at andres.cruz@utexas.edu and Matt at mjmartin@utexas.edu.

1 R and RStudio

2 Tidyverse basics I

3 Matrices

3.1 Introduction

3.1.1 Scalars

- One number (12, for example) is referred to as a scalar.
- This can be thought of as a 1x1 matrix.

3.1.2 Vectors

- We can put several scalars together to make a vector.
- An example is:

$$\begin{bmatrix} 12 \\ 14 \\ 15 \end{bmatrix} = b$$

- Since this is a column of numbers, we cleverly refer to it as a column vector.
- Another example is:

$$\begin{bmatrix} 12 & 14 & 15 \end{bmatrix} = d$$

- This is called a row vector.

3.2 Operators

3.2.1 Summation

- Recall the summation operator $\sum$, which lets us perform an operation on a sequence of numbers (often but not always a vector)

$$x = \begin{bmatrix} 12 & 7 & -2 & 0 & 1 \end{bmatrix}$$

- We can then calculate...

$$\sum_{i=1}^{3} x_i$$

- Which is...

$$12 + 7 + -2 = 17$$

3.2.2 Product

- Recall the product operator $\prod$, which can also perform operations over a sequence of numbers

$$z = \begin{bmatrix} 5 & -3 & 5 & 1 \end{bmatrix}$$

- We can then calculate...

$$\prod_{i=1}^{4} z_i$$

- Which multiplies out to...

$$5 * -3 * 5 * 1 = -75$$

3.3 Matrices

3.3.1 Basics

- We can append vectors together to form a matrix:

$$\begin{bmatrix} 12 & 14 & 15 \\ 115 & 22 & 127 \\ 193 & 29 & 219 \end{bmatrix} = A$$

- We always refer to the dimensions of matrices by row then column (R x C).
  - Find a way to remember that knowledge *permanently*.
- So, A is a $3x3$ matrix.
  - Note that matrices are usually designated by capital letters, and sometimes **bolded**, too.

3.3.2 Structure

- How do we refer to specific elements of the matrix?
- Matrix A is an $m \times n$ matrix where $m = n = 3$
- More generally, matrix B is an $m \times n$ matrix where the elements look like this:

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \dots & b_{2n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ b_{m1} & b_{m2} & b_{m3} & \dots & b_{mn} \end{bmatrix}$$

- Thus b_{23} refers to the second unit down and third across.

3.4 Matrix operations

3.4.1 Addition and subtraction

- Addition and subtraction are logical.
- Matrices have *exactly* the same dimensions for these operations.
- Add or subtract each element with the corresponding element from the other matrix:

$$A \pm B = C$$

$$c_{ij} = a_{ij} \pm b_{ij} \; \forall i, j$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \pm \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$
$$=$$
$$\begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & a_{13} \pm b_{13} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & a_{23} \pm b_{23} \\ a_{31} \pm b_{31} & a_{32} \pm b_{32} & a_{33} \pm b_{33} \end{bmatrix}$$

—

Practice

$$A = \begin{bmatrix} 1 & 4 & 2 \\ -2 & -1 & 0 \\ 0 & -1 & 3 \end{bmatrix}$$

$$B = \begin{bmatrix} 5 & 1 & 0 \\ 2 & -1 & 0 \\ 7 & 1 & 2 \end{bmatrix}$$

Calculate $A + B$

Practice

$$A = \begin{bmatrix} 6 & -2 & 8 & 12 \\ 4 & 42 & 8 & -6 \\ -14 & 5 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 18 & 42 & 3 & 7 \\ 0 & -42 & 15 & 4 \\ -7 & 0 & 21 & -18 \end{bmatrix}$$

Calculate $A - B$

3.4.2 Scalar multiplication

- Recall that a scalar is a single number.
- Multiply each value in the matrix by the scalar.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$cA = \begin{bmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \\ ca_{31} & ca_{32} & ca_{33} \end{bmatrix}$$

—

Practice

$$A = \begin{bmatrix} 1 & 4 & 2 \\ 8 & -1 & 3 \\ 0 & -2 & 3 \end{bmatrix}$$

$$B = \begin{bmatrix} -15 & 1 & 5 \\ 2 & -42 & 0 \\ 7 & 1 & 6 \end{bmatrix}$$

Find $2 \times A$ and $-3 \times B$

3.4.3 Matrix multiplication

- Requirement: the two matrices must be *conformable*.
- This means that the number of columns in the first matrix equals the number of rows in the second.
- When multiplying $A \times B$, if A is $m \times n$, B must have n rows.
- The resulting matrix will have the number of rows in the first, and the number of columns in the second.
- For example, if A is $i \times k$ and B is $k \times j$, then $A \times B$ will be $i \times j$.

Which of the following can we multiply? What will be the dimensions of the resulting matrix?

$$b = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 1 \end{bmatrix} \quad M = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 2 & 4 \\ 2 & 3 & 2 \end{bmatrix} \quad L = \begin{bmatrix} 6 & 5 & -1 \\ 1 & 4 & 3 \end{bmatrix}$$

Why can't we multiply in the opposite order?

- Multiply each row by each column, summing up each pair of multiplied terms
- The element in position ij is the sum of the products of elements in the ith row of the first matrix (A) and the corresponding elements in the jth column of the second matrix (B).

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

- Let's do some examples on the board.

3.4.4 Properties

- Addition and subtraction:
  - Associative: $(A \pm B) \pm C = A \pm (B \pm C)$
  - Communicative $A \pm B = B \pm A$
- Multiplication:
  - $AB \neq BA$
  - $A(BC) = (AB)C$
  - $A(B + C) = AB + AC$
  - $(A + B)C = AC + BC$

4 Tidyverse basics II

5 Functions and loops

6 Calculus

6.1 Theory

- Calculus is about dealing with infinitesimal values.
- We are going to focus on two big ideas:
  - Derivatives
  - Integrals

6.1.1 Derviative

- "Derivative" is just a fancy term for slope.
- Slope is the rate of change $\frac{\delta y}{\delta x}$ or $\frac{dy}{dx}$.
- Specifically, the derivative is the *instantaneous* rate of change.
- We need slope for our statistics, which are all about fitting lines.
- We also need slope for taking maxima and minima.
- The equation for a line is $y = mx + b$. What is its slope?

6.1.2 Calculating derivatives

- Slope is rise over run, which is $\dfrac{f(x+\Delta x) - f(x)}{\Delta x}$
- To see why, consider the slope of a line connecting two points:

$$m = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

- We can define $x_2 = x_1 + \Delta x$ (or equivalently $\Delta x = x_2 - x_1$)

$$m = \frac{f(x_1 + \Delta x) - f(x_1)}{\Delta x}$$

- As we've seen, for a curve, we need to be infinitely close for our line's defining points, yielding
$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$
- This gives us this instantaneous slope (rate of change) of a function at every point on its domain. The above equation is the definition of the derivative.

6.1.3 Notation

- $\frac{d}{dx}f(x)$ is read "The derivative of f of x with respect to x." - You can also say "The instantaneous rate of change in f of x with respect to x."
- If $y = f(x)$, $\frac{dy}{dx}$ is "The derivative of y with respect to x". - Warning: Do not try to cancel out the d's, no matter how tempting it is.
- There is the advantage of always specifying the variable with respect to which we're differentiating (it's the one in the denominator).

Lagrange's prime notation: - $f'(x)$ (read: "f prime x") is the derivative of $f(x)$. - This is useful when it is clear which variable were are referring to (e.g., when there's only one).

- What is $\dfrac{d(x^2)}{dx}$? - x^2 - $2x^{2-1}$ -$2x$
- What is $\frac{d(4x^3)}{dx}$?
  - $4x^3$
  - $4 * 3x^{3-1}$
  - $12x^2$

Practice

Take the derivative of each of these.

$$x^3$$

$$3x^2$$

$$60x^{11}$$

$$x$$

$$\frac{4}{x^2}$$

$$9\sqrt{x}$$

$$6x^{5/2}$$

$$11,596,232$$

Evaluate the derivatives at $x = 2$ and $x = -1$

$$x^3$$

$$3x^2$$

$$60x^{11}$$

$$x$$

$$\frac{4}{x^2}$$

$$9\sqrt{x}$$

$$6x^{5/2}$$

$$11,596,232$$

Practice

Take the derivative of each of these.

$$x^3$$

$$3x^2$$

$$60x^{11}$$

$$x$$

$$\frac{4}{x^2}$$

$$9\sqrt{x}$$

$$6x^{5/2}$$

$$11,596,232$$

Evaluate the derivatives at $x = 2$ and $x = -1$

$$x^3$$

$$3x^2$$

$$60x^{11}$$

$$\frac{x}{4}$$
$$x^2$$

$$9\sqrt{x}$$

$$6x^{5/2}$$

$$11,596,232$$

6.1.4 Special functions

A few functions have particular rules:

- $\frac{d(ln(x))}{dx} = \frac{1}{x}$
- $\frac{d(log_b(x))}{dx} = \frac{1}{x * ln(b)}$
- $\frac{d(e^x)}{dx} = e^x$
- $\frac{d(a^x)}{dx} = a^x ln(a)$
- $\frac{dy}{dx}c = 0$
- $\frac{d(x^x)}{dx} = x^x(1 + ln(x))$

6.1.5 Derivatives with addition and substraction

- Easiest rule to remember:

$$\frac{d(f(x) \pm g(x))}{dx} = f'(x) \pm g'(x)$$

Practice

Take the derivative of each of these

$$x^2 + x + 5$$

$$x^4 - 4x^3 + 5x^2 + 8x - 6$$

$$3x^5 - 6x^2$$

$$5x^2 + 8\sqrt{x} - \frac{1}{x}$$

$$ln(x) + 5e^x - 4x^3$$

6.2 Advanced rules

6.2.1 Product rule

- A little more complicated:

$$\frac{d(f(x) \times g(x))}{dx} = f'(x)g(x) + g'(x)f(x)$$

- Example: $2x \times 3x$

Practice

Take the derivative of each of these:

$$x^3 * x$$

$$e^x * x^2$$

$$ln(x) * x^{-3}$$

Remember, $\frac{d(f(x)*g(x))}{dx} = f'(x)g(x) + g'(x)f(x)$.

6.2.2 Quotient rule

$$\frac{d\frac{f(x)}{g(x)}}{dx} = \frac{f'(x)g(x) - g'(x)f(x)}{[g(x)]^2}$$

If you're having trouble with this, just apply the product rule to:

$$\frac{d[f(x) * g^{-1}(x)]}{dx}$$

Remember, $\frac{d\frac{f(x)}{g(x)}}{dx} = \frac{f'(x)g(x) - g'(x)f(x)}{[g(x)]^2}$.

6.2.3 Chain rule

$$\frac{d[f(g(x))]}{dx} = f'(g(x)) * g'(x)$$

Let's take the derivative of a function of a function:

$$\frac{d[ln(x^2)]}{dx}$$

$$f(x) = ln(x), g(x) = x^2$$

$$f'(x) = \frac{1}{x}, g'(x) = 2x$$

$$\frac{1}{x^2} * 2x = \frac{2}{x}$$

Practice

Take the derivative of each of these:

$$(3x^4 - 8)^2$$

$$e^{x^2}$$

Remember, $\frac{d(f(g(x))}{dx} = f'(g(x)) * g'(x)$.

6.2.4 Second derivative

- Same process as taking single derivative, except input for second derivative is output from first.
- Second derivative tells us whether the slope of a function is increasing, decreasing, or staying the same at any point x on the function's domain.
- Example: driving a car.
  - $f(x)$ = distance traveled at time x
  - $f'(x)$ = speed at time x
  - $f''(x)$ = acceleration at time x

Graph $f(x) = x^2$, $f'(x)$, and $f''(x)$.

$$\frac{d^2(x^4)}{dx^2} = f''(x^4)$$

- First, we take the first derivative:

$$f'(x^4) = 4x^3$$

- Then we use that output to take the second derivative:

$$f''(x^4) = f'(4x^3) = 12x^2$$

Practice

Take the second derivative of the following functions:

$$x^5$$

$$6x^2$$

$$4ln(x)$$

$$3x$$

$$4x^{3/2}$$

7 Probability

8 Simulations

9 Text analysis

10 Wrap up

References