Mathematics Prerequisites

Contents

1 Mathematics

> Math is not just a way of calculating numerical answers; it is a way of thinking, using clear definitions for concepts and rigorous logic to organize our thoughts and back up our assertions.

Cheng (2025)

These lecture notes use:

- algebra
- precalculus
- univariate calculus
- linear algebra
- vector calculus

Some key results are listed here.

1.1 Elementary Algebra

Mastery of Elementary Algebra[1] (a.k.a. "College Algebra") is a prerequisite for calculus, which is a prerequisite for Epi 202 and Epi 203, which are prerequisites for this course (Epi 204). Nevertheless, each year, some Epi 204 students are still uncomfortable with algebraic manipulations of mathematical formulas. Therefore, I include this section as a quick reference.

1.1.1 Equalities

Theorem 1.1 (Equalities are transitive)**.** *If $a = b$ and $b = c$, then $a = c$*

Theorem 1.2 (Substituting equivalent expressions)**.** *If $a = b$, then for any function $f(x)$, $f(a) = f(b)$*

1.1.2 Inequalities

Theorem 1.3. *If $a < b$, then $a + c < b + c$*

Theorem 1.4 (negating both sides of an inequality)**.** *If $a < b$, then: $-a > -b$*

Theorem 1.5. *If $a < b$ and $c \geq 0$, then $ca < cb$.*

Theorem 1.6.

$$-a = (-1) * a$$

1.1.3 Sums

Theorem 1.7 (adding zero changes nothing)**.**

$$a + 0 = a$$

Theorem 1.8 (Sums are symmetric)**.**

$$a + b = b + a$$

Theorem 1.9 (Sums are associative)**.**

When summing three or more terms, the order in which you sum them does not matter:

$$(a + b) + c = a + (b + c)$$

---

[1] https://en.wikipedia.org/wiki/Elementary_algebra

1.1.4 Products

Theorem 1.10 (Multiplying by 1 changes nothing)**.**

$$a \times 1 = a$$

Theorem 1.11 (Products are symmetric)**.**

$$a \times b = b \times a$$

Theorem 1.12 (Products are associative)**.**

$$(a \times b) \times c = a \times (b \times c)$$

1.1.5 Division

Theorem 1.13 (Division can be written as a product)**.**

$$\frac{a}{b} = a \times \frac{1}{b}$$

1.1.6 Sums and products together

Theorem 1.14 (Multiplication is distributive)**.**

$$a(b + c) = ab + ac$$

1.1.7 Quotients

Definition 1.1 (Quotients, fractions, rates)**.**

A **quotient**, **fraction**, or **rate** is a division of one quantity by another:

$$\frac{a}{b}$$

In epidemiology, rates typically have a quantity involving time or population in the denominator.

c.f. https://en.wikipedia.org/wiki/Rate_(mathematics)

Definition 1.2 (Ratios)**.** A **ratio** is a quotient in which the numerator and denominator are measured using the same unit scales.

c.f. https://en.wikipedia.org/wiki/Ratio

Definition 1.3 (Proportion)**.** In statistics, a "proportion" typically means a ratio where the numerator represents a subset of the denominator.

See https://en.wikipedia.org/wiki/Population_proportion.

See also https://en.wikipedia.org/wiki/Proportion_(mathematics) for other meanings.

Definition 1.4 (Proportional)**.** Two functions $f(x)$ and $g(x)$ are **proportional** if their ratio $\frac{f(x)}{g(x)}$ does not depend on x. (c.f. https://en.wikipedia.org/wiki/Proportionality_(mathematics))

Additional reference for elementary algebra: https://en.wikipedia.org/wiki/Population_proportion#Mathematical_definition

---

1.2 Exponentials and Logarithms

Theorem 1.15 (logarithm of a product is the sum of the logs of the factors)**.**

$$\log a \cdot b = \log a + \log b$$

Corollary 1.1 (logarithm of a quotient)**.**

The logarithm of a quotient is equal to the difference of the logs of the factors:

$$\log \frac{a}{b} = \log a - \log b$$

Theorem 1.16 (logarithm of an exponential function)**.**

$$log\{a^b\} = b \cdot log\{a\}$$

Theorem 1.17 (exponential of a sum)**.**

The exponential of a sum is equal to the product of the exponentials of the addends:

$$exp\{a+b\} = exp\{a\} \cdot exp\{b\}$$

Corollary 1.2 (exponential of a difference)**.**

The exponential of a difference is equal to the quotient of the exponentials of the addends:

$$exp\{a-b\} = \frac{exp\{a\}}{exp\{b\}}$$

Theorem 1.18 (exponential of a product)**.**

$$a^{bc} = \left(a^b\right)^c = \left(a^c\right)^b$$

Corollary 1.3 (natural exponential of a product)**.**

$$exp\{ab\} = (exp\{a\})^b = (exp\{b\})^a$$

Exercise 1.1. For $a \geq 0,\ b, c \in \mathbb{R}$, When does $(a^b)^c = a^{(b^c)}$?

Solution 1.1. Short answer: rarely (that's all you need to know for this course).

Long answer:

If $(a^b)^c = a^{(b^c)}$, then since $(a^b)^c = a^{bc}$, we have:

$$\begin{aligned} a^{bc} &= a^{(b^c)} \\ \log\{a^{bc}\} &= \log\left\{a^{(b^c)}\right\} \\ bc \cdot \log\{a\} &= b^c \cdot \log\{a\} \end{aligned} \tag{1}$$

Equation 1 holds in each of the following cases:

1. $bc = b^c$ (see Exercise 1.2).
2. $a = 1$ (i.e., $\log\{a\} = 0$).
3. $a = 0$ (i.e., $\log\{a\} = -\infty$) *and* $\text{sign}\{bc\} = \text{sign}\{b^c\}$.

In particular, when $a = 0$ and $c = 0$, $bc = 0$ and $b^c = 1$ (for any $b \in \mathbb{R}$), so $\text{sign}\{bc\} \neq \text{sign}\{b^c\}$, and $(a^b)^c \neq a^{(b^c)}$:

$$\begin{aligned}(a^b)^c &= (0^b)^0 \\ &= 1\end{aligned}$$

$$\begin{aligned}a^{(b^c)} &= 0^{(b^0)} \\ &= 0^1 \\ &= 0\end{aligned}$$

Exercise 1.2. For $b, c \in \mathbb{R}$, when does $b^c = bc$?

Solution 1.2. $bc = b^c$ in each of the following cases:

1. $c = 1$.
2. $b = 0$ and $c > 0$.
3. $b = \exp\left\{\frac{\log c}{c-1}\right\}$ (for $c \geq 0$).

See the red contours in Figure 2 for a visualization.

```r
`b*c_f` <- function(b, c) b*c
`b^c_f` <- function(b, c) b^c
values_b <- seq(0, 5, by = .01)
values_c <- seq(-.5, 3, by = .01)

`b*c` <- outer(values_b, values_c, `b*c_f`)
`b^c` <- outer(values_b, values_c, `b^c_f`)
`b^c`[is.infinite(`b^c`)] = NA

opacity <- .3
z_min <- min(`b*c`, `b^c`, na.rm = TRUE)
z_max <- 5
plotly::plot_ly(
  x = ~values_b,
  y = ~values_c
) |>
  plotly::add_surface(
    z = ~ t(`b*c`),
    contours = list(
      z = list(
        show = TRUE,
        start = -1,
        end = 1,
        size = .1
      )
    ),
    name = "b*c",
    showscale = FALSE,
    opacity = opacity,
    colorscale = list(c(0, 1), c("green", "green"))
  ) |>
  plotly::add_surface(
    opacity = opacity,
    colorscale = list(c(0, 1), c("red", "red")),
    z = ~ t(`b^c`),
    contours = list(
      z = list(
        show = TRUE,
        start = z_min,
        end = z_max,
        size = .2
      )
    ),
    showscale = FALSE,
    name = "b^c"
  ) |>
  plotly::layout(
    scene = list(
      xaxis = list(
        # type = "log",
        title = "b"
      ),
      yaxis = list(
        # type = "log",
        title = "c"
      ),
      zaxis = list(
        # type = "log",
        range = c(z_min, z_max),
        title = "outcome"
      ),
      camera = list(eye = list(x = -1.25, y = -1.25, z = 0.5)),
      aspectratio = list(x = .9, y = .8, z = 0.7)
```

```r
`b^c - b*c_f` <- function(b, c) `b^c_f`(b,c) - `b*c_f`(b,c)

mat1 <- outer(values_b, values_c, `b^c - b*c_f`)
mat1[is.infinite(mat1)] = NA

opacity <- .3
plotly::plot_ly(
  x = ~values_b,
  y = ~values_c
) |>
  plotly::add_surface(
    z = ~ t(mat1),
    contours = list(
      z = list(
        show = TRUE,
        start = 0,
        end = 1,
        size = 1,
        color = "red"
      )
    ),
    name = "b^c - b*c",
    showscale = TRUE,
    opacity = opacity
  ) |>
  plotly::layout(
    scene = list(
      xaxis = list(
        # type = "log",
        title = "b"
      ),
      yaxis = list(
        # type = "log",
        title = "c"
      ),
      zaxis = list(
        title = "outcome"
      ),
      camera = list(eye = list(x = -1.25, y = -1.25, z = 0.5)),
      aspectratio = list(x = .9, y = .8, z = 0.7)
    )
  )
```

Theorem 1.19 (exp{} and log{} are mutual inverses)**.**

$$exp\{log\{a\}\} = log\{exp\{a\}\} = a$$

2 Derivatives

Theorem 2.1 (Constant rule)**.**

$$\frac{\partial}{\partial x} c = 0$$

Theorem 2.2 (Power rule)**.** *If a is constant with respect to x, then:*

$$\frac{\partial}{\partial x} ay = a\frac{\partial x}{\partial y}$$

Theorem 2.3 (Power rule)**.**

$$\frac{\partial}{\partial x} x^q = qx^{q-1}$$

Theorem 2.4 (Derivative of natural logarithm)**.**

$$log'\{x\} = \frac{1}{x} = x^{-1}$$

Theorem 2.5 (derivative of exponential)**.**

$$exp'\{x\} = exp\{x\}$$

Theorem 2.6 (Product rule)**.**

$$(ab)' = ab' + ba'$$

Theorem 2.7 (Quotient rule)**.**

$$(a/b)' = a'/b - (a/b^2)b'$$

Theorem 2.8 (Chain rule)**.**

$$\begin{aligned}\frac{\partial a}{\partial c} &= \frac{\partial a}{\partial b}\frac{\partial b}{\partial c} \\ &= \frac{\partial b}{\partial c}\frac{\partial a}{\partial b}\end{aligned}$$

or in Euler/Lagrange notation[2]*:*

$$(f(g(x)))' = g'(x)f'(g(x))$$

Corollary 2.1 (Chain rule for logarithms)**.**

$$\frac{\partial}{\partial x} \log f(x) = \frac{f'(x)}{f(x)}$$

Proof. Apply Theorem 2.8 and Theorem 2.4. ☐

---

[2] https://en.wikipedia.org/wiki/Notation_for_differentiation#Lagrange's_notation

3 Linear Algebra

Definition 3.1 (Dot product/linear combination/inner product)**.** For any two real-valued vectors $\tilde{x} = (x_1, \dots, x_n)$ and $\tilde{y} = (y_1, \dots, y_n)$, the **dot-product**, **linear combination**, or **inner product** of $\tilde{x}$ and $\tilde{y}$ is:

$$\tilde{x} \cdot \tilde{y} = \tilde{x}^\top \tilde{y} \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_i y_i$$

Note

See also the definitions in

- Dobson and Barnett (2018), §1.3 (equation 1.1, page 7)
- Kaplan (2022), here[a].
- wikipedia[b]

"Linear combination" can also refer to weighted sums of vectors, or in other words matrix-vector multiplication.

The dot-product has a different generalization for two matrices; see wikipedia[c] for more.

[a] https://www.mosaic-web.org/MOSAIC-Calculus/Textbook/Linear-combinations/28-Vectors.html#geometry-arithmetic

[b] https://en.wikipedia.org/wiki/Linear_combination

[c] https://en.wikipedia.org/wiki/Dot_product#Dyadics_and_matrices

---

Theorem 3.1 (Dot product is symmetric)**.** *The dot product is symmetric:*

$$\tilde{x} \cdot \tilde{y} = \tilde{y} \cdot \tilde{x}$$

Proof. Apply:

- Definition 3.1
- symmetry of scalar multiplication
- Definition 3.1 again

□

4 Vector Calculus

(adapted from Fieller (2016), §7.2[3])

Let $\tilde{x}$ and $\tilde{\beta}$ be vectors of length p, or in other words, matrices of length $p \times 1$:

$$\tilde{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

$$\tilde{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

[3] https://www.taylorfrancis.com/chapters/mono/10.1201/9781315370200-7/vector-matrix-calculus-nick-fieller?context=ubx&refId=c310b723-786a-4f33-ae56-720a6cccd3a1

Definition 4.1 (Transpose)**.** The transpose of a row vector is the column vector with the same sequence of entries:

$$\tilde{x}' \equiv \tilde{x}^\top \equiv [x_1, x_2, ..., x_p]$$

Example 4.1 (Dot product as matrix multiplication)**.**

$$\begin{aligned}\tilde{x} \cdot \tilde{\beta} &= \tilde{x}^\top \tilde{\beta} \\ &= [x_1, x_2, ..., x_p] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \\ &= x_1\beta_1 + x_2\beta_2 + ... + x_p\beta_p\end{aligned}$$

Theorem 4.1 (Transpose of a sum)**.**

$$(\tilde{x} + \tilde{y})^\top = \tilde{x}^\top + \tilde{y}^\top$$

Definition 4.2 (Vector derivative)**.** If $f(\tilde{\beta})$ is a function that takes a vector $\tilde{\beta}$ as input, such as $f(\tilde{\beta}) = x'\tilde{\beta}$, then:

$$\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta}) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} f(\tilde{\beta}) \\ \frac{\partial}{\partial \beta_2} f(\tilde{\beta}) \\ \vdots \\ \frac{\partial}{\partial \beta_p} f(\tilde{\beta}) \end{bmatrix}$$

Definition 4.3 (Row-vector derivative)**.** If $f(\tilde{\beta})$ is a function that takes a vector $\tilde{\beta}$ as input, such as $f(\tilde{\beta}) = x'\tilde{\beta}$, then:

$$\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta}) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} f(\tilde{\beta}) & \frac{\partial}{\partial \beta_2} f(\tilde{\beta}) & \cdots & \frac{\partial}{\partial \beta_p} f(\tilde{\beta}) \end{bmatrix}$$

Theorem 4.2 (Row and column derivatives are transposes)**.**

$$\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta}) = \left(\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta}) \right)^\top$$

$$\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta}) = \left(\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta}) \right)^\top$$

Theorem 4.3 (Derivative of a dot product)**.**

$$\frac{\partial}{\partial \tilde{\beta}} \tilde{x} \cdot \tilde{\beta} = \frac{\partial}{\partial \tilde{\beta}} \tilde{\beta} \cdot \tilde{x} = \tilde{x}$$

This looks a lot like non-vector calculus, except that you have to transpose the coefficient.

Proof.

$$\frac{\partial}{\partial \beta}(x^\top \beta) = \begin{bmatrix} \frac{\partial}{\partial \beta_1}(x_1\beta_1 + x_2\beta_2 + ... + x_p\beta_p) \\ \frac{\partial}{\partial \beta_2}(x_1\beta_1 + x_2\beta_2 + ... + x_p\beta_p) \\ \vdots \\ \frac{\partial}{\partial \beta_p}(x_1\beta_1 + x_2\beta_2 + ... + x_p\beta_p) \end{bmatrix}$$
$$= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$
$$= \tilde{x}$$

□

Definition 4.4 (Quadratic form)**.** A **quadratic form** is a mathematical expression with the structure

$$\tilde{x}^\top \mathbf{S} \tilde{x}$$

where $\tilde{x}$ is a vector and $\mathbf{S}$ is a matrix with compatible dimensions for vector-matrix multiplication.

Quadratic forms occur frequently in regression models. They are the matrix-vector generalizations of the scalar quadratic form $cx^2 = xcx$.

Theorem 4.4 (Derivative of a quadratic form)**.** *If S is a $p \times p$ matrix that is constant with respect to β, then:*

$$\frac{\partial}{\partial \beta} \beta' S \beta = 2 S \beta$$

This is like taking the derivative of cx^2 with respect to x in non-vector calculus.

Corollary 4.1 (Derivative of a simple quadratic form)**.**

$$\frac{\partial}{\partial \tilde{\beta}} \tilde{\beta}' \tilde{\beta} = 2\tilde{\beta}$$

This is like taking the derivative of x^2.

Theorem 4.5 (Vector chain rule)**.**

$$\frac{\partial z}{\partial \tilde{x}} = \frac{\partial y}{\partial \tilde{x}} \frac{\partial z}{\partial y}$$

or in Euler/Lagrange notation:

$$(f(g(\tilde{x})))' = \tilde{g}'(\tilde{x}) f(g(\tilde{x}))$$

See https://quickfem.com/finite-element-analysis/, specifically https://quickfem.com/wp-content/uploads/IFEM.AppF_.pdf

See also https://en.wikipedia.org/wiki/Gradient#Relationship_with_Fr%C3%A9chet_derivative

This chain rule is like the univariate chain rule (Theorem 2.8), but the order matters now. The version presented here is for the gradient[4] (column vector); the total derivative[5] (row vector) would be the transpose of the gradient[6].

Corollary 4.2 (Vector chain rule for quadratic forms)**.**

$$\frac{\partial}{\partial \tilde{\beta}}\left(\tilde{\varepsilon}(\tilde{\beta}) \cdot \tilde{\varepsilon}(\tilde{\beta})\right) = \left(\frac{\partial}{\partial \tilde{\beta}}\tilde{\varepsilon}(\tilde{\beta})\right)\left(2\tilde{\varepsilon}(\tilde{\beta})\right)$$

5 Additional resources

5.1 Calculus

- Kaplan (2022)
- Khuri (2003)
- Banner (2007)
- Miller (2016)
  - http://www.youtube.com/watch?v=xYzQL0TUtBA
  - http://www.youtube.com/watch?v=Ps2SBo_WjoE

5.2 Linear Algebra and Vector Calculus

- Fieller (2016)
- Banerjee and Roy (2014)
- Searle and Khuri (2017)

5.3 Numerical Analysis

- Hua Zhou[7]'s lecture notes for "UCLA Biostat 216 - Mathematical Methods for Biostatistics" (2023 Fall)[8]

5.4 Real Analysis

- Grinberg (2017)

Banerjee, Sudipto, and Anindya Roy. 2014. *Linear Algebra and Matrix Analysis for Statistics*. Vol. 181. Crc Press Boca Raton. https://www.routledge.com/Linear-Algebra-and-Matrix-Analysis-for-Statistics/Banerjee-Roy/p/book/9781420095388.

Banner, Adrian D. 2007. *The Calculus Lifesaver : All the Tools You Need to Excel at Calculus*. A Princeton Lifesaver Study Guide. Princeton University Press. https://press.princeton.edu/books/paperback/9780691130880/the-calculus-lifesaver.

Cheng, Eugenia. 2025. "Opinion | How Math Turned Me from a D.E.I. Skeptic to a Supporter." *The New York Times*. https://www.nytimes.com/2025/09/05/opinion/math-dei.html.

Dobson, Annette J, and Adrian G Barnett. 2018. *An Introduction to Generalized Linear Models*. 4th ed. CRC press. https://doi.org/10.1201/9781315182780.

Fieller, Nick. 2016. *Basics of Matrix Algebra for Statistics with R*. Chapman; Hall/CRC. https://doi.org/10.1201/9781315370200.

[4]https://en.wikipedia.org/wiki/Gradient
[5]https://en.wikipedia.org/wiki/Total_derivative
[6]https://en.wikipedia.org/wiki/Gradient#Relationship_with_total_derivative
[7]https://hua-zhou.github.io/
[8]https://ucla-biostat-216.github.io/2023fall/schedule/schedule.html

Grinberg, Raffi. 2017. *The Real Analysis Lifesaver: All the Tools You Need to Understand Proofs*. 1st ed. Princeton Lifesaver Study Guides. Princeton University Press. https://press.princeton.edu/books/paperback/9780691172934/the-real-analysis-lifesaver.

Kaplan, Daniel. 2022. *MOSAIC Calculus*. Www.mosaic-web.org. www.mosaic-web.org[9].

Khuri, André I. 2003. *Advanced Calculus with Applications in Statistics*. John Wiley & Sons. https://www.wiley.com/en-us/Advanced+Calculus+with+Applications+in+Statistics%2C+2nd+Edition-p-9780471391043.

Miller, Steven J. 2016. *The Probability Lifesaver: Calculus Review Problems*. https://web.williams.edu/Mathematics/sjmiller/public_html/probabilitylifesaver/index.htm#:~:text=http%3A//web.williams.edu/Mathematics/sjmiller/public_html/probabilitylifesaver/supplementalchap_calcreview.pdf.

Searle, Shayle R, and Andre I Khuri. 2017. *Matrix Algebra Useful for Statistics*. John Wiley & Sons.

[9] https://www.mosaic-web.org