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).

$$a + 0 = a$$

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

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:

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

1.1.4 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

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:

$$a^{bc} = a^{(b^c)}$$

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

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

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.

```
`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)),
```

```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[a]*:*

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

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

---

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

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

Proof

Proof. Apply Theorem 2.8 and Theorem 2.4. □

---

3 Linear Algebra

3.1 Vectors

Definition 3.1 (Column vector)**.** A **column vector** of length p is an ordered list of p numbers, written vertically:

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

Column vectors are the default convention in these notes and in most statistics textbooks. They are also called $p \times 1$ *matrices*.

Definition 3.2 (Transpose)**.** The **transpose** of a column vector $\tilde{x}$ is the **row vector** with the same sequence of entries, written horizontally:

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

The transpose operation converts a column vector to a row vector, or more generally, swaps the rows and columns of a matrix (Definition 3.10).

3.1.1 Special vectors

Definition 3.3 (Zero vector)**.** The **zero vector** $\tilde{0}$ of length p has all entries equal to zero:

$$\tilde{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The zero vector is the additive identity for vector addition: $\tilde{x} + \tilde{0} = \tilde{x}$ for any vector $\tilde{x}$ of the same length.

Definition 3.4 (Ones vector)**.** The **ones vector** $\tilde{1}$ of length p has all entries equal to one:

$$\tilde{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

The dot product $\tilde{1}^\top \tilde{x} = \tilde{1} \cdot \tilde{x} = \sum_{i=1}^p x_i$ is the sum of all entries of $\tilde{x}$.

Definition 3.5 (Indicator vector / standard basis vector)**.** The j-th **indicator vector** (or **standard basis vector**) $\tilde{e}_j$ of length p has a 1 in position j and 0s elsewhere:

$$(\tilde{e}_j)_i = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad \tilde{e}_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \text{position } j$$

They are also called **unit vectors** or **standard basis vectors**.

Theorem 3.1 (Indicator vectors select entries)**.** *For any vector $\tilde{x}$ of length p and any $j \in \{1, \dots, p\}$:*

$$\tilde{e}_j^\top \tilde{x} = x_j$$

Proof

Proof. Writing the product componentwise:

$$\begin{aligned} \tilde{e}_j^\top \tilde{x} &= \sum_{i=1}^p (\tilde{e}_j)_i \, x_i \\ &= \sum_{i=1}^p \begin{cases} 1 \cdot x_i & \text{if } i = j \\ 0 \cdot x_i & \text{if } i \neq j \end{cases} \\ &= x_j \end{aligned}$$

□

Definition 3.6 (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 3.2 (Dot product is symmetric)**.** *The dot product is symmetric:*

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

Proof

Proof. Apply:

- Definition 3.6
- symmetry of scalar multiplication
- Definition 3.6 again

□

Example 3.1 (Dot product as matrix multiplication)**.** The dot product of two column vectors $\tilde{x}$ and $\tilde{\beta}$ can be written as a matrix product of the row vector $\tilde{x}^\top$ with the column vector $\tilde{\beta}$:

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

3.1.2 Orthogonality

Definition 3.7 (Orthogonal vectors)**.** Two vectors $\tilde{x}$ and $\tilde{y}$ of the same length are **orthogonal** (written $\tilde{x} \perp \tilde{y}$) if their dot product is zero:

$$\tilde{x} \perp \tilde{y} \iff \tilde{x}^\top \tilde{y} = 0$$

Orthogonality generalizes the geometric notion of perpendicularity to arbitrary dimensions.

Definition 3.8 (Orthonormal vectors)**.** A set of vectors $\{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_k\}$ is **orthonormal** if the vectors are mutually orthogonal and each has unit length:

$$\tilde{x}_i^\top \tilde{x}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The indicator vectors $\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_p$ (Definition 3.5) form an orthonormal set.

3.2 Matrices

Definition 3.9 (Matrix)**.** A **matrix** of dimensions $m \times n$ is a rectangular array of $m \cdot n$ numbers, arranged in m rows and n columns:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

The entry in row i and column j is denoted a_{ij} or $(\mathbf{A})_{ij}$. A column vector of length p is a special case: a $p \times 1$ matrix. A row vector of length p is a $1 \times p$ matrix.

3.2.1 Matrix transpose

Definition 3.10 (Matrix transpose)**.** The **transpose** of an $m \times n$ matrix $\mathbf{A}$ is the $n \times m$ matrix $\mathbf{A}^\top$ obtained by swapping the rows and columns of $\mathbf{A}$:

$$(\mathbf{A}^\top)_{ij} = a_{ji}$$

Theorem 3.3 (Transpose of a sum)**.**

$$(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$$

In particular, for column vectors $\tilde{x}$ and $\tilde{y}$:

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

Theorem 3.4 (Transpose of a product)**.** *For compatible matrices $\mathbf{A}$ and $\mathbf{B}$:*

$$(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top \mathbf{A}^\top$$

The order of the factors reverses when transposing a product.

3.2.2 Matrix addition

Definition 3.11 (Zero matrix)**.** The $m \times n$ **zero matrix** $\mathbf{0}_{m\times n}$ (or $\mathbf{0}$ when dimensions are clear from context) has all entries equal to zero:

$$\mathbf{0}_{m\times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Definition 3.12 (Matrix addition)**.** Two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimensions $m \times n$ can be added element-wise:

$$(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij}$$

Theorem 3.5 (Matrix addition is commutative)**.**

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

Theorem 3.6 (Matrix addition is associative)**.**

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$$

Theorem 3.7 (Zero matrix is the additive identity)**.**

$$\mathbf{A} + \mathbf{0} = \mathbf{A}$$

Theorem 3.8 (Additive inverse)**.** *For any matrix $\mathbf{A}$, the matrix $-\mathbf{A}$ (defined by $(-\mathbf{A})_{ij} = -a_{ij}$) satisfies:*

$$\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$$

3.2.3 Scalar multiplication

Definition 3.13 (Scalar multiplication)**.** A matrix $\mathbf{A}$ can be multiplied by a scalar c:

$$(c\mathbf{A})_{ij} = c \cdot a_{ij}$$

3.2.4 Matrix multiplication

Definition 3.14 (Matrix multiplication)**.** The **product** of an $m \times k$ matrix $\mathbf{A}$ and a $k \times n$ matrix $\mathbf{B}$ is the $m \times n$ matrix $\mathbf{C} = \mathbf{AB}$ with entries:

$$c_{ij} = \sum_{s=1}^{k} a_{is}\, b_{sj}$$

Matrix multiplication is only defined when the number of columns in $\mathbf{A}$ equals the number of rows in $\mathbf{B}$.

Matrix multiplication is **not** commutative in general: $\mathbf{AB} \neq \mathbf{BA}$.

Theorem 3.9 (Matrix multiplication is associative)**.**

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

Theorem 3.10 (Matrix multiplication is distributive over addition)**.**

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$$

3.2.5 Matrix-vector multiplication

Definition 3.15 (Matrix-vector multiplication)**.** The product of an $m \times p$ matrix $\mathbf{A}$ and a $p \times 1$ column vector $\tilde{x}$ is the $m \times 1$ column vector $\mathbf{A}\tilde{x}$ with entries:

$$(\mathbf{A}\tilde{x})_i = \sum_{j=1}^{p} a_{ij}\, x_j$$

Matrix-vector multiplication is a generalization of the dot product. Each entry of the result is a dot product of a row of $\mathbf{A}$ with the vector $\tilde{x}$.

3.3 Special Matrices

See also Definition 3.11 for the zero matrix.

Definition 3.16 (Square matrix)**.** A matrix is **square** if it has the same number of rows as columns. The number of rows (= columns) is the **order** of the matrix.

Definition 3.17 (Matrix power)**.** For a square matrix $\mathbf{A}$ of order p and a positive integer k, the k-th **power** of $\mathbf{A}$ is:

$$\mathbf{A}^k = \underbrace{\mathbf{A}\,\mathbf{A} \cdots \mathbf{A}}_{k \text{ copies}}$$

In particular, $\mathbf{A}^2 = \mathbf{A}\mathbf{A}$.

Definition 3.18 (Identity matrix)**.** The $p \times p$ **identity matrix** $\mathbf{I}_p$ (or $\mathbf{I}$ when the size is clear from context) has ones on the main diagonal and zeros elsewhere:

$$(\mathbf{I}_p)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad \mathbf{I}_p = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Theorem 3.11 (Identity matrix is a multiplicative identity)**.** *For any $m \times p$ matrix $\mathbf{A}$:*

$$\mathbf{A}\,\mathbf{I}_p = \mathbf{A}$$

$$\mathbf{I}_m\,\mathbf{A} = \mathbf{A}$$

Definition 3.19 (Symmetric matrix)**.** A square matrix $\mathbf{A}$ is **symmetric** if $\mathbf{A}^\top = \mathbf{A}$, i.e., $a_{ij} = a_{ji}$ for all i and j.

Covariance matrices and information matrices are symmetric.

Definition 3.20 (Diagonal matrix)**.** A square matrix $\mathbf{D}$ is a **diagonal matrix** if all off-diagonal entries are zero: $d_{ij} = 0$ whenever $i \neq j$:

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}$$

Diagonal matrices are denoted $\mathbf{D} = \operatorname{diag}(d_1, d_2, \ldots, d_p)$, where $d_1, \ldots, d_p$ are the diagonal entries.

Definition 3.21 (Matrix inverse)**.** For a square $p \times p$ matrix $\mathbf{A}$, the **inverse** $\mathbf{A}^{-1}$ (if it exists) is the unique matrix satisfying:

$$\mathbf{A}\,\mathbf{A}^{-1} = \mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}_p$$

A matrix that has an inverse is called **invertible** or **non-singular**.

Theorem 3.12 (Inverse of a product)**.** *For invertible matrices $\mathbf{A}$ and $\mathbf{B}$:*

$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

Definition 3.22 (Idempotent matrix)**.** A square matrix $\mathbf{A}$ is **idempotent** if

$$\mathbf{A}^2 = \mathbf{A}$$

Definition 3.23 (Projection matrix)**.** A square matrix $\mathbf{P}$ is a **projection matrix** (also called an **orthogonal projector**) if it is both symmetric and idempotent:

$$\mathbf{P}^\top = \mathbf{P} \qquad \text{and} \qquad \mathbf{P}^2 = \mathbf{P}$$

Theorem 3.13 (Complement of a projection matrix)**.** *If $\mathbf{P}$ is a projection matrix, then $\mathbf{I}-\mathbf{P}$ is also a projection matrix.*

Proof

Proof. We verify symmetry and idempotency.
Symmetry:

$$(\mathbf{I}-\mathbf{P})^\top = \mathbf{I}^\top - \mathbf{P}^\top = \mathbf{I}-\mathbf{P}$$

Idempotency:

$$\begin{aligned}(\mathbf{I}-\mathbf{P})^2 &= (\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})\\ &= \mathbf{I}-\mathbf{P}-\mathbf{P}+\mathbf{P}^2\\ &= \mathbf{I}-\mathbf{P}-\mathbf{P}+\mathbf{P}\\ &= \mathbf{I}-\mathbf{P}\end{aligned}$$

□

Theorem 3.14 (Hat matrix is a projection matrix)**.** *In a linear regression model with full-rank design matrix $\mathbf{X}$, the* ***hat matrix***

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$$

is a projection matrix.

Proof

Proof. We verify symmetry and idempotency.
Symmetry:

$$\begin{aligned}\mathbf{H}^\top &= \left(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\right)^\top\\ &= \left(\mathbf{X}^\top\right)^\top \cdot \left((\mathbf{X}^\top\mathbf{X})^{-1}\right)^\top \cdot \mathbf{X}^\top\\ &= \mathbf{X}\cdot(\mathbf{X}^\top\mathbf{X})^{-1}\cdot\mathbf{X}^\top\\ &= \mathbf{H}\end{aligned}$$

where the third line uses $\left(\mathbf{X}^\top\right)^\top = \mathbf{X}$ and the fact that $\mathbf{X}^\top\mathbf{X}$ is symmetric, so its inverse is also symmetric ($\left((\mathbf{X}^\top\mathbf{X})^{-1}\right)^\top = (\mathbf{X}^\top\mathbf{X})^{-1}$).

Idempotency:

$$\begin{aligned}\mathbf{H}^2 &= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \cdot \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \\ &= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{X}^\top\mathbf{X})(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \\ &= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \\ &= \mathbf{H}\end{aligned}$$

□

The hat matrix appears in the formula for fitted values in linear regression: $\hat{\tilde{y}} = \mathbf{X}\hat{\tilde{\beta}} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\tilde{y} = \mathbf{H}\tilde{y}$. It "puts a hat" on $\tilde{y}$ — hence the name.

Theorem 3.15 (Projection matrices produce orthogonal decompositions)**.** *If* $\mathbf{P}$ *is a projection matrix and* $\tilde{v}$ *is any vector of compatible dimension, then the two components of the decomposition*

$$\tilde{v} = \underbrace{\mathbf{P}\tilde{v}}_{projected} + \underbrace{(\mathbf{I}-\mathbf{P})\tilde{v}}_{residual}$$

are orthogonal:

$$\mathbf{P}\tilde{v} \perp (\mathbf{I}-\mathbf{P})\tilde{v}$$

Proof

Proof.

$$\begin{aligned}(\mathbf{P}\tilde{v})^\top(\mathbf{I}-\mathbf{P})\tilde{v} &= \tilde{v}^\top\mathbf{P}^\top(\mathbf{I}-\mathbf{P})\tilde{v} \\ &= \tilde{v}^\top\mathbf{P}(\mathbf{I}-\mathbf{P})\tilde{v} \\ &= \tilde{v}^\top(\mathbf{P}-\mathbf{P}^2)\tilde{v} \\ &= \tilde{v}^\top(\mathbf{P}-\mathbf{P})\tilde{v} \\ &= \tilde{v}^\top\mathbf{0}\,\tilde{v} \\ &= 0\end{aligned}$$

where the second line uses symmetry ($\mathbf{P}^\top = \mathbf{P}$) and the fourth line uses idempotency ($\mathbf{P}^2 = \mathbf{P}$). □

3.4 Quadratic Forms

Definition 3.24 (Quadratic form)**.** A **quadratic form** is a mathematical expression of the structure

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

where $\tilde{x}$ is a $p \times 1$ vector and $\mathbf{S}$ is a $p \times p$ matrix.

Quadratic forms are the matrix generalizations of the scalar expression cx^2. They occur frequently in statistics:

- The residual sum of squares in linear regression
  (1) is a quadratic form.
- The variance of a linear combination of estimates
  (2) is a quadratic form: $\mathrm{Var}\left(\tilde{x}^\top\hat{\tilde{\beta}}\right) = \tilde{x}^\top\,\mathrm{Var}\left(\hat{\tilde{\beta}}\right)\tilde{x}$.

Theorem 3.16 (Symmetric part of a quadratic form)**.** *If* $\mathbf{S}$ *is a square matrix, then*

$$\tilde{x}^\top \mathbf{S}\tilde{x} = \tilde{x}^\top \left(\frac{1}{2}(\mathbf{S} + \mathbf{S}^\top)\right)\tilde{x}.$$

So the value of a quadratic form depends only on the symmetric part of $\mathbf{S}$*.*

3.5 Design Matrix

Definition 3.25 (Design matrix)**.** In a regression model with n observations and p predictors, the **design matrix** (or **model matrix**) $\mathbf{X}$ is the $n \times p$ matrix whose i-th row is the covariate vector $\tilde{x}_i^\top$ for observation i:

$$\mathbf{X} = \begin{bmatrix} \tilde{x}_1^\top \\ \tilde{x}_2^\top \\ \vdots \\ \tilde{x}_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

The product $\mathbf{X}\tilde{\beta}$ collects all the linear predictors $\tilde{x}_i^\top\tilde{\beta}$ into a single $n \times 1$ vector:

$$\mathbf{X}\tilde{\beta} = \begin{bmatrix} \tilde{x}_1^\top\tilde{\beta} \\ \vdots \\ \tilde{x}_n^\top\tilde{\beta} \end{bmatrix}$$

The matrix $\mathbf{X}^\top\mathbf{X}$ is a $p \times p$ symmetric matrix that appears in the OLS estimator $\hat{\tilde{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\tilde{y}$.

4 Vector Calculus

(adapted from Fieller (2016), §7.2[2])

This section covers derivatives of functions of vectors and matrices. Linear algebra prerequisites — including vectors, matrices, transpose, dot product, and quadratic forms — are covered in Section 3.

Let $\tilde{x}$ and $\tilde{\beta}$ be column vectors of length p (see Definition 3.1 and Definition 3.6).

Definition 4.1 (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 4.2 (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:

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

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

Theorem 4.1 (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 4.2 (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

Proof.

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

□

Theorem 4.3 (Derivative of a quadratic form)**.** *For a quadratic form (Definition 3.24), if S is a $p \times p$ matrix that is constant with respect to β, then:*

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

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 4.4 (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[3] (column vector); the total derivative[4] (row vector) would be the transpose of the gradient[5].

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

---

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

5.4 Real Analysis

- Grinberg (2017)

References

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.

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[8].

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.

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