Math Review

PHS Launch 2022

Rienna Russo

August 24, 2022

Outline

Welcome!

Welcome to the **math review**! Today we will be covering some key concepts that will be helpful in building intuition for the methods we learn about in PHS 2000.

Some reminders:

- We do not expect you to memorize all of the information presented today.
- These slides can hopefully serve as a reference that you can return to throughout the course.
- Today may be a refresher for some folks while for others today may an introduction.
- Everyone comes to this program with a unique background, which is what makes PHS so special!

Welcome!

We do not want you feeling like this at the end of today’s session…

So please feel free to stop me at any time and ask questions!

Linear Algebra Basics

Quick Reminder: Vectors and Matrices

Vector: a structured set of inputs (e.g., numbers) arranged in a list

$$\mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

Matrix: a structured set of vectors all with the same length (or *dimensionality*)

$$\mathrm{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p}$$

Transpose

We can flip a vector on its side by transposing it.

$$\mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \qquad \mathrm{y}^T = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}_{1\times n}$$

We can do the same with matrices. The former columns (variables) are now rows. The former rows (participants) are now columns.

$$\mathrm{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n\times p} \qquad \mathrm{X}^T = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{pn} \end{bmatrix}_{p\times n}$$

Vector Multiplication

We can only multiply vectors of the same length and the inner dimensions must match.

If we start with two vectors a and b with length p, we can transpose one of our vectors and then multiply.

$$\mathrm{a}^{\mathrm{T}}\mathrm{b} = \begin{bmatrix} a_1 & a_2 & \dots & a_p \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = a_1 b_1 + a_2 b_2 + \ldots + a_p b_p = \sum_{i=1}^{p} a_i b_i$$

Vector Multiplication: Example

We have two vectors, a and b, of length 3.

$$\mathrm{a}^{\mathrm{T}} = \begin{bmatrix} 4 & 2 & 5 \end{bmatrix}_{1\times 3} \quad \mathrm{b} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}_{3\times 1}$$

$$\mathrm{a}^{\mathrm{T}}\mathrm{b} = (4 \times 1) + (2 \times 3) + (5 \times 2)$$

$$\mathrm{a}^{\mathrm{T}}\mathrm{b} = 20$$

Matrix Multiplication

As with vectors, we can only multiply matrices that have the same inner dimensions.

$$\mathrm{C}_{m\times n}\mathrm{D}_{n\times p} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix} \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1p} \\ d_{21} & d_{22} & \cdots & d_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{np} \end{bmatrix}$$

The product of two original matrices $\mathrm{C}_{m\times n}\mathrm{D}_{n\times p}$ is a new matrix, $\mathrm{E}_{m\times p}$, that retains the original matrices outer dimensions.

$$\mathrm{E}_{m\times p} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{mp} \end{bmatrix}$$

Matrix Multiplication

The first element of our new matrix, $\mathrm{E}_{m\times p}$, is the product of the first row from $\mathrm{C}_{m\times n}$ and the first column from $\mathrm{D}_{n\times p}$

$$\mathrm{e}_{11} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \end{bmatrix} \begin{bmatrix} d_{11} \\ d_{21} \\ d_{31} \\ \vdots \\ d_{n1} \end{bmatrix} = \textstyle\sum_{i=1}^{n} c_i d_i$$

Matrix Multiplication: Example

We have two matrices, $\mathrm{C}_{2\times 3}$ and $\mathrm{D}_{3\times 2}$.

$$\mathrm{C}_{2\times 3}\mathrm{D}_{3\times 2} = \begin{bmatrix} 4 & 2 & 5 \\ 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 2 & 2 \end{bmatrix}$$

$$\mathrm{E}_{2\times 2} = \begin{bmatrix} (4\times 1)+(2\times 3)+(5\times 2) & (4\times 2)+(2\times 4)+(5\times 2) \\ (3\times 1)+(4\times 3)+(1\times 2) & (3\times 2)+(4\times 4)+(1\times 2) \end{bmatrix}$$

$$\mathrm{E}_{2\times 2} = \begin{bmatrix} 20 & 26 \\ 17 & 24 \end{bmatrix}$$

Matrix Multiplication: Example

Integer multiplication is **communicative**, i.e., changing the order of the elements does not affect the results: $3 \times 2 = 2 \times 3$
Matrix multiplication is **not communicative**: $\mathrm{C}_{2\times3}\mathrm{D}_{3\times2} \neq \mathrm{D}_{3\times2}\mathrm{C}_{2\times3}$.

$$\mathrm{D}_{3\times2}\mathrm{C}_{2\times3} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 4 & 2 & 5 \\ 3 & 4 & 1 \end{bmatrix}$$

$$\mathrm{F}_{3\times3} = \begin{bmatrix} (1\times4)+(2\times3) & (1\times2)+(2\times4) & (1\times5)+)2\times1) \\ (3\times4)+(4\times3) & (3\times2)+(4\times4) & (3\times5)+(4\times1) \\ (2\times4)+(2\times3) & (2\times2)+(2\times4) & (2\times5)+(2\times1) \end{bmatrix}$$

Matrix Multiplication: Example

$$\mathrm{F}_{3\times 3} = \begin{bmatrix} (1\times 4)+(2\times 3) & (1\times 2)+(2\times 4) & (1\times 5)+)2\times 1) \\ (3\times 4)+(4\times 3) & (3\times 2)+(4\times 4) & (3\times 5)+(4\times 1) \\ (2\times 4)+(2\times 3) & (2\times 2)+(2\times 4) & (2\times 5)+(2\times 1) \end{bmatrix}$$

$$\mathrm{F}_{3\times 3} = \begin{bmatrix} 10 & 10 & 7 \\ 24 & 22 & 19 \\ 14 & 12 & 12 \end{bmatrix}$$

Order matters!

Questions?

Design Matrix

A design matrix is identical to your dataset, except there is an additional columns of 1's.

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p+1}$$

The design matrix has dimensionality of $n \times (p + 1)$
i.e. the number of observations or rows in your dataset $\times$ the number of columns in your dataset.

Connection to Course Concepts

A sneak peak at linear regression...

We can multiply our design matrix with dimensions $n \times (p+1)$ by a transpose vector of β coefficients:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

Identity Matrix

The product of a matrix and the appropriate identity matrix will return the original matrix.

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$XI = X$$

Similar to how the product of a number and 1 gives you the original number: $a \times 1 = a$!

Inverse Matrix

The product of a matrix and its inverse matrix will return the identity matrix.

$$\mathrm{X}\mathrm{X}^{-1} = \mathrm{I}$$

Similar to how the inverse of a number multiplied by the original number gives you 1*:* $a \times \frac{1}{a} = 1$ *!*

Questions?

Functions

Function Definition

A procedure that relates each element x of a set X (the domain) to exactly one element y of another set Y (the range).

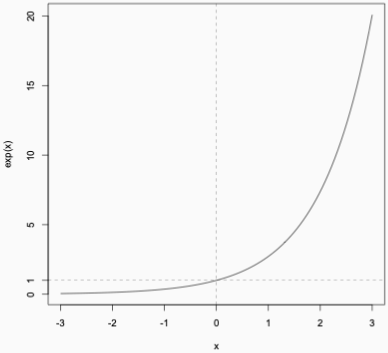
Each input from set X has only a single output, i.e., every X uniquely produces a particular Y.

Some examples:

- $f(x)$ arbitrary function
- $\mu(x)$ function for a mean
- $\pi(x)$ function for a probability
- $exp(x)$ natural exponential function

Natural Exponential Function

The exponential function maps any one-dimensional real number to some other non-zero, positive one-dimensional real number.



$$y = exp(x) = e^x$$

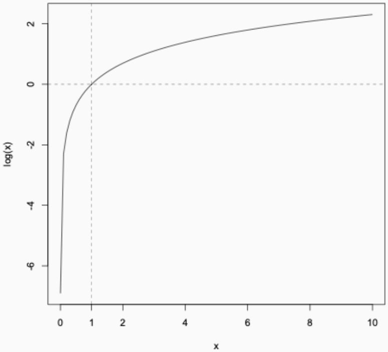
NOTE: e stands for Euler's constant: $e \approx 2.718$

Domain: $(-\infty, +\infty)$

Range: $[0, +\infty]$

Natural Logarithm Function

The log function maps any one-dimensional non-zero, positive real number to some other one-dimensional real number.



$$y = log(x)$$

NOTE: When you see $log(x)$ in this class, it refers to the natural base e log or ln

Domain: $[0, +\infty)$

Range: $(-\infty, +\infty)$

The $log(x)$ function is the inverse of the $exp(x)$ function.
$log(x) = exp^{-1}(x)$ and $exp(log(x)) = x$

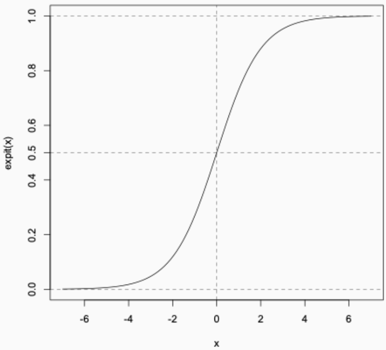
Laws of Exponents and Logarithms

Law of Exponents	Law of Logarithms
$exp(a+b) = exp(a) \times exp(b)$	$log(a) + log(b) = log(a \times b)$
$exp(a-b) = \frac{exp(a)}{exp(b)}$	$log(a) - log(b) = log(\frac{a}{b})$
$exp(a)^b = exp(a \times b)$	$log(a^b) = b \times log(a)$

Questions?

Expit Function

The expit function maps any one-dimensional real number to some other one-dimensional real number between 0 and 1.



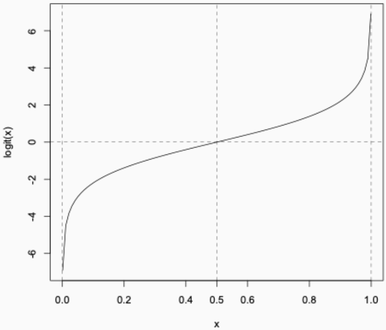
$$y = expit(x) = \frac{exp(x)}{(1+exp(x))}$$

Domain: $(-\infty, +\infty)$

Range: $[0, 1]$

Logit Function

The logit function maps any one-dimensional real number between 0 and 1 to some other one-dimensional real number.



$y = logit(x)$

$= log(odds(x)) = log(\frac{x}{(1-x)})$

Domain: $[0, 1]$

Range: $(-\infty, +\infty)$

The $logit(x)$ function is the inverse of the $expit(x)$ function.

$logit(x) = expit^{-1}(x)$ and $logit(expit(x)) = x$

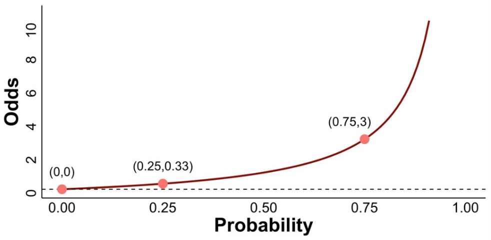
A Brief Review of Probability and Odds

Probability The number of successes, x, out of the total number of trials, n.

- Proportion
- Range: $[0, 1]$
- Formula: $\frac{x}{n}$

Odds The number of successes, x, compared to the number of failures, $n - x$.

- Ratio
- Range: $[0, +\infty)$
- Formula: $\frac{x}{(n-x)}$



Note: As probabilities get larger and larger, the difference between probability and odds increases.

Questions?

Linear Regression

The linear regression function maps a $p+1$-dimensional vector of real numbers ($x_i = x_0, x_1, x_2, ..., x_p$) to some one-dimensional real number (y_i) using a set of unspecified parameters $\beta_0, \beta_1, ..., \beta_p$

$$f(x) = y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$$

This can also be written as $y_i = \mathrm{x_i} \beta^{\mathrm{T}}$.

Domain: $(-\infty, +\infty)$

Range: $(-\infty, +\infty)$

Linear Regression

For a population, the linear regression function maps an $n \times (p+1)$ matrix (i.e., your design matrix) to an n-dimensional vector (y)

$$f(x) = y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$$

$$y = \mathrm{X}\beta^{\mathrm{T}}$$

Connection to Course Concepts

We can multiply our design matrix with dimensions $n \times (p+1)$ by a transpose vector of β coefficients:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

Questions?

Putting it all together

Data Example

Your friend tells you they are thinking about applying for the popular reality show, The Bachelor, and wants to know more about their chances of receiving a rose on the first night. You decide to model the probability of receiving a rose on the first night as a function of height, age, and number of compliments.

Data Example

Data about 10 participants are stored in matrix X with the following 4 columns:

- Height (continuous)
- Age (continuous)
- Number of compliments (count)
- Received a rose (binary)

$$\mathrm{X} = \begin{bmatrix} 73 & 26 & 4 & 0 \\ 67 & 31 & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 64 & 28 & 3 & 1 \end{bmatrix}$$

Data Example

Data about 10 participants are stored in matrix X with the following 4 columns:

- Height (continuous)
- Age (continuous)
- Number of compliments (count)
- Received a rose (binary)

$$\mathrm{X} = \begin{bmatrix} 73 & 26 & 4 & 0 \\ 67 & 31 & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 64 & 28 & 3 & 1 \end{bmatrix}$$

What is the dimensionality of this matrix?

Data Example

By adding a column of 1s, we can create a design matrix for our dataset.

$$X = \begin{bmatrix} 1 & 73 & 26 & 4 & 0 \\ 1 & 67 & 31 & 0 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 64 & 28 & 3 & 1 \end{bmatrix}$$

What is the new dimensionality of this matrix?

Data Example

We decide to model the probability of receiving a rose as a function of height, age, and number of compliments.

We define the following model and variables:

$$\pi(x) = y = \beta_0 + \beta_1 height + \beta_2 age + \beta_3 compliments$$

- π: probability of receiving a rose
- x: vector of variables $(height, age, compliments)$
- $height$: height in inches (continuous)
- age: age in years (continuous)
- $compliments$: compliments given on the first night (count)

Data Example

We decide to model the probability of getting a rose as a function of height, age, and number of compliments.

We define the following model and variables:

$$\pi(x) = y = \beta_0 + \beta_1 height + \beta_2 age + \beta_3 compliments$$

- π: probability of receiving a rose
- x: vector of variables $(height, age, compliments)$
- $height$: height in inches (continuous)
- age: age in years (continuous)
- $compliments$: compliments given on the first night (count)

Any issues?

Problem

$$\pi(x) = y = \beta_0 + \beta_1 height + \beta_2 age + \beta_3 compliments$$

Since x and $\beta_0, \beta_1, \beta_2,$ andβ_3 can be any real numbers, then y could take on any real number.

But we want the probability, which needs to be constrained to $[0, 1]$.

Problem

$$\pi(x) = y = \beta_0 + \beta_1 height + \beta_2 age + \beta_3 compliments$$

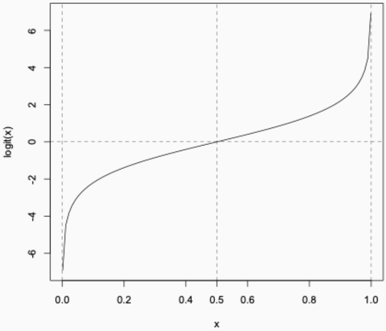
Since x and $\beta_0, \beta_1, \beta_2,$ andβ_3 can be any real numbers, then y could take on any real number.

But we want the probability, which needs to be constrained to $[0, 1]$.

What can we do to this equation?

Recall the Logit Function

The logit function maps any one-dimensional real number between 0 and 1 to some other one-dimensional real number.



$y = logit(x)$

$= log(odds(x)) = log(\frac{x}{(1-x)})$

Domain: $[0, 1]$

Range: $(-\infty, +\infty)$

Solution

If we apply the **logit** function to the left side, our outcome, π, will take values between $[0, 1]$, while x and β can be any real number.

$$logit(\pi(x)) = \beta_0 + \beta_1 height + \beta_2 age + \beta_3 compliments$$

Solution

We can re-write this model like this:

$$log(\frac{\pi(x)}{1-\pi(x)}) = \beta_0 + \beta_1 height + \beta_2 age + \beta_3 compliments$$

Solution

We can re-write this model like this:

$$log(\frac{\pi(x)}{1-\pi(x)}) = \beta_0 + \beta_1 height + \beta_2 age + \beta_3 compliments$$

What does this model now look like?

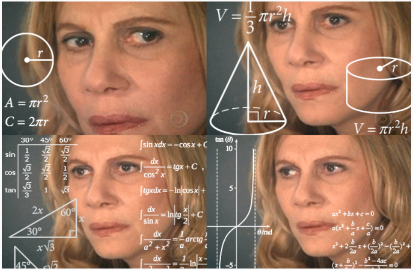
Connection to Course Concepts

Generalized linear models are models for a transformation of the expected outcome as a linear function of the model parameters.

E.g., a **link** function, such as the log or the $logit$ functions can remove restrictions on the range of the outcome variable

Additional Resources

In case you're feeling like this right now



Khan Academy Links

- Vectors
- Matrix Transformations
- Exponential and Logarithm Functions
- Linear Regression

PHS 2000A Teaching Team

- **TFs:** Rienna Russo, Anna Siefkas, Matt Lee, and Sudipta Saha
- **Instructors:** Jarvis Chen and Issa Dahabreh

Your peers!

Thank you!

Many thanks to past **PHS 2000A Teaching Fellows** who taught this material:

Emma McGee
Kat Sadikova
Louisa Smith
Aaron Sarvet
Gabe Schwartz
Bryan Patenaude