Acknowledgement of Country

- The University of Queensland (UQ) acknowledges the Traditional Owners and their custodianship of the lands on which we meet.
- We pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.
- We recognise their valuable contributions to Australian and global society.

Image: Digital reproduction of *A guidance through time* by Casey Coolwell and Kyra Mancktelow

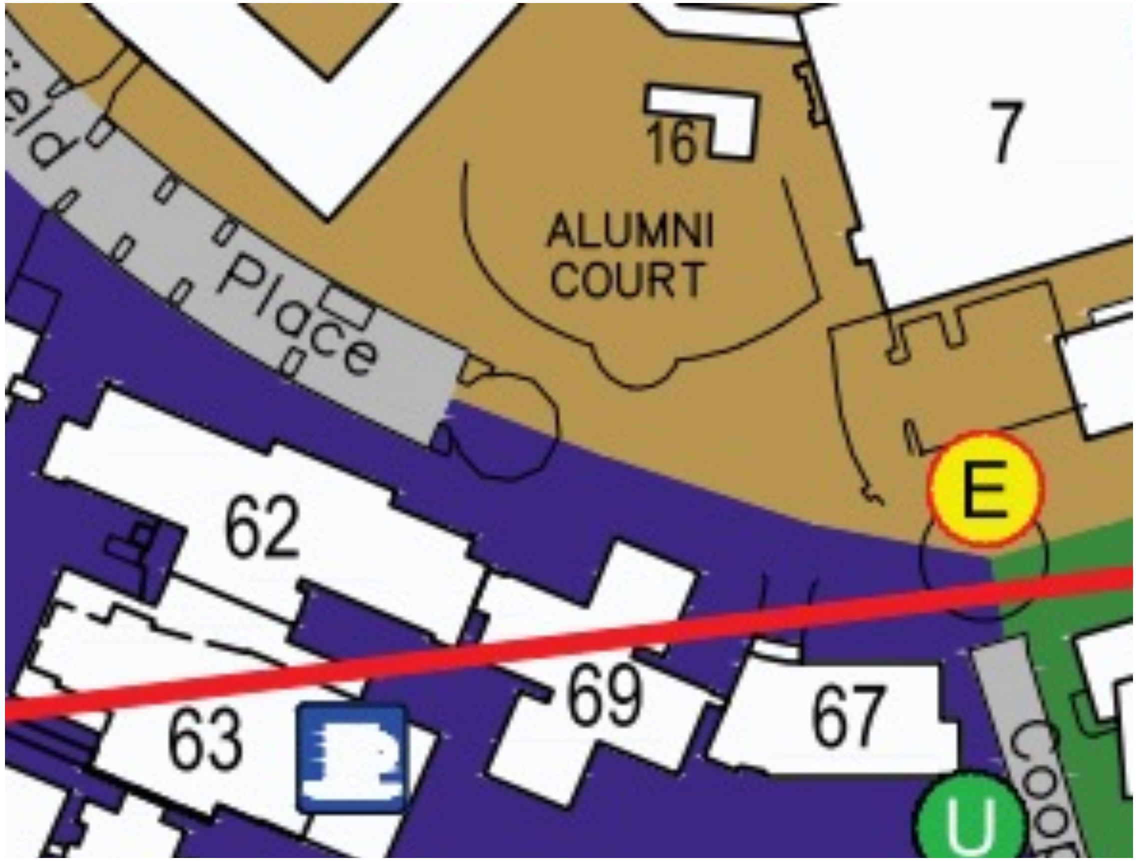
General Information:

- We are currently located in Building 69

Emergency evacuation point

- Food court and bathrooms are in Building 63

- If you are experiencing cold/flu symptoms or have had COVID in the last 7 days please ensure you are wearing a mask for the duration of the module



Data Agreement

To maximize your learning experience, we will be working with genuine human genetic data, during this module.

Access to this data requires agreement to the following in to comply with human genetic data ethics regulations

Please email pctgadmin@imb.com.au with your name and the below statement to confirm that you agree with the following:

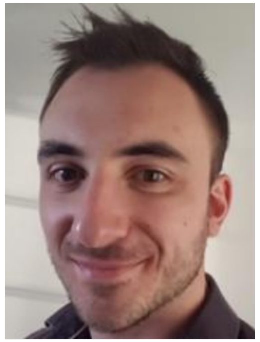
“I agree that access to data is provided for educational purposes only and that I will not make any copy of the data outside the provided computing accounts.”

Plan for today

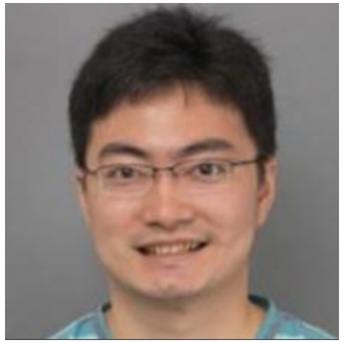
9-12pm	Basic Stats – Kathryn Kemper 3xlecture 3xpractical
12-1pm	Lunch break
1-4:30pm	GWAS – Allan McRae 2xlecture 2xpractical
4:30pm	Walk to IMB for dinner @5pm (for registered participants)

Allan

Kath



Valentine



Jian

2022 Winter School

Introduction to matrix algebra

Kathryn Kemper

Matrix Algebra

- Why?
  - Store many numbers using a single symbol
  - Compact
  - Many known results written in matrix form

Introduction to Matrix Algebra: Outline

- Why matrix algebra is important

- Definition of a matrix & some special types of matrices
- Matrix operations; multiplication by scalar, addition/subtraction, multiplication, inverse, transpose

- Intro to PCA - Principal Component Analysis railway car

What is a matrix?

A rectangular array of numbers set in rows and columns,

e.g. **B** is a 2x3 matrix

$$\boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

OR

$$\boldsymbol{B} = \begin{bmatrix} 4 & 1 & 3 \\ 5 & -1 & 2 \end{bmatrix}$$

b_{ij} is the *ij* element of the matrix

Vector is a matrix with either 1 row or 1 column, e.g.

$\boldsymbol{w} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$ is a 2x1 column vector

OR

$\boldsymbol{v} = \begin{bmatrix} 3 & 5 \end{bmatrix}$ is a 1x2 row vector

Special types of matrices

- Square matrix

  (equal number of rows & columns)

$$G = \begin{bmatrix} 1 & 3 & 7 \\ 9 & 4 & 3 \\ 2 & 1 & 2 \end{bmatrix}$$

off-diagonal elements

diagonal elements

- Diagonal matrix

  (sq. matrix with 0's for all off-diagonal elements)

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

- Identity matrix

  (diagonal matrix with 1's for all diagonal elements)

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

Special types of matrices

- ‘J’ matrix

A vector of 1’s, written **1** (or sometimes $\mathbf{1_n}$)

- Symmetric matrix

A square matrix where element *ij* equals element *ji*, e.g.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 7 \\ 3 & 4 & -3 \\ 7 & -3 & 2 \end{bmatrix}$$

i.e.,

$a_{12} = a_{21} = 3$

$a_{13} = a_{31} = 7$

$a_{23} = a_{32} = -3$

Special types of matrices, partitioned matrices

- A matrix that has been broken into sections or ‘blocks’
- For convince / shorthand

e.g.

$$E = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where:

- A is a n x 1 matrix
- B is a n x p matrix
- C is a q x 1 matrix
- D is a q x p matrix

...and E is therefore a (n+q) x (1+p) matrix

Matrix operations: multiplication by scalar

- A ‘scalar’ is a 1x1 matrix, or an ordinary number
- multiply each element of matrix by the scalar

$$\boldsymbol{2B} = 2\begin{bmatrix} 4 & 1 & 3 \\ 5 & -1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 4x2 & 1x2 & 3x2 \\ 5x2 & -1x2 & 2x2 \end{bmatrix}$$

$$= \begin{bmatrix} 8 & 2 & 6 \\ 10 & -2 & 4 \end{bmatrix}$$

Matrix operations: Addition or subtraction

- Only possible if matrices have same number of rows and columns (i.e. they are of the same order and are 'conformable')

Let $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{Y}$,
with $\boldsymbol{X}$ and $\boldsymbol{Y}$ being 2x2 matrices of $\boldsymbol{X} = \begin{bmatrix} \mathbf{3} & \mathbf{5} \\ \mathbf{-1} & \mathbf{2} \end{bmatrix}$ and $\boldsymbol{Y} = \begin{bmatrix} \mathbf{1} & \mathbf{7} \\ \mathbf{2} & \mathbf{9} \end{bmatrix}$

$$\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{x_{11}} + \boldsymbol{y_{11}} & \boldsymbol{x_{12}} + \boldsymbol{y_{12}} \\ \boldsymbol{x_{21}} + \boldsymbol{y_{21}} & \boldsymbol{x_{22}} + \boldsymbol{y_{22}} \end{bmatrix}$$

$$\boldsymbol{W} = \begin{bmatrix} \mathbf{3+1} & \mathbf{5+7} \\ \mathbf{-1+2} & \mathbf{2+9} \end{bmatrix} = \begin{bmatrix} \mathbf{4} & \mathbf{9} \\ \mathbf{1} & \mathbf{11} \end{bmatrix}$$

Matrix operations: Transpose

- The transpose of matrix $\boldsymbol{B}$ is usually written $\boldsymbol{B}'$ (or sometimes $\boldsymbol{B}^{\boldsymbol{T}}$)
- The transpose of $\boldsymbol{B}$ is the matrix whose element *ji* are equal to the *ij* element of $\boldsymbol{B}$ i.e. $b_{ij} = b'_{ji}$
  - ➢Turn the columns of $\boldsymbol{B}$ into the rows of $\boldsymbol{B}'$

If $\boldsymbol{B} = \begin{bmatrix} 4 & 1 & 3 \\ 5 & -1 & 2 \end{bmatrix}$ then $\boldsymbol{B}' = \begin{bmatrix} 4 & 5 \\ 1 & -1 \\ 3 & 2 \end{bmatrix}$

Note that $\boldsymbol{B} \neq \boldsymbol{B}'$ (unless $\boldsymbol{B}$ is symmetric)

Matrix operations: multiplication

- Order is important!
- Only possible if number of columns 1st matrix = number of rows 2nd matrix

If $\mathbf{A}$ has order r *x* c & $\mathbf{B}$ has order c *x* n

- then $\mathbf{AB}$ exists with order r *x* n
- The ij element of $\mathbf{AB}$ is 'sum product' of row i from matrix $\mathbf{A}$ and column j from $\boldsymbol{B}$

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 5 \end{bmatrix}, B = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

$$AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix} = \begin{bmatrix} 1x2 + 2x1 & 1x3 + 2x4 \\ 2x2 + 3x1 & 2x3 + 3x4 \\ 0x2 + 5x1 & 0x3 + 5x4 \end{bmatrix} = \begin{bmatrix} 4 & 11 \\ 7 & 18 \\ 5 & 20 \end{bmatrix}$$

Matrix operations: multiplication

- Order is important!
- If $\mathbf{A}$ has order r *x* c & $\mathbf{B}$ has order c *x* n
- then $\mathbf{AB}$ exists with order r *x* n
  - $\mathbf{BA}$ doesn't exist unless r = n
  - Even if $\mathbf{BA}$ exists it isn't necessarily equal to $\mathbf{AB}$

- e.g. if $\mathbf{A}$ is a 3x2 and $\boldsymbol{B}$ is a 2x3 then
  - $\mathbf{AB}$ is a 3x3 matrix
  - $\mathbf{BA}$ is a 2x2 matrix

Matrix operations: multiplication

- Why define it this way?
- It's useful

e.g. to store simultaneous equations

2x + 3y = 15

x - y = 1

can be rewritten as $\mathbf{Av} = \mathbf{b}$; where $\mathbf{A} = \begin{bmatrix} \mathbf{2} & \mathbf{3} \\ \mathbf{1} & \mathbf{-1} \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}$ & $\mathbf{b} = \begin{bmatrix} \mathbf{15} \\ \mathbf{1} \end{bmatrix}$

Matrix operations: multiplication

- The identity matrix (matrix with 1's on diagonal) has a special property,
  - ➢Any matrix multiplied by its identity matrix returns the original matrix

- If $\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix}$, the $\mathbf{AI} = \mathbf{A}$

Let's try it: $\begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2x1 + 3x0 & 2x0 + 3x1 \\ 1x1 - 1x0 & 1x0 - 1x1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix}$

Matrix operations: inverse

- Some (not all!) square matrices have an inverse
- The inverse of $\mathbf{A}$ is written as $\boldsymbol{A}^{-\mathbf{1}}$
- The inverse of a matrix is one where when multiplied by the original matrix, it returns the identity matrix
  - $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
- Relatively easy to calculate for 2x2 matrix:
  - If $\mathbf{A} = \begin{bmatrix} \boldsymbol{a}_{11} & \boldsymbol{a}_{12} \\ \boldsymbol{a}_{21} & \boldsymbol{a}_{22} \end{bmatrix}$ then
  - $\boldsymbol{A}^{-1} = \frac{1}{|A|}\begin{bmatrix} \boldsymbol{a}_{22} & -\boldsymbol{a}_{12} \\ -\boldsymbol{a}_{21} & \boldsymbol{a}_{11} \end{bmatrix}$ where |A| is the determinate of A, and
  - |A| = $(\boldsymbol{a}_{11}\boldsymbol{a}_{22} - \boldsymbol{a}_{12}\boldsymbol{a}_{21})$

Matrix operations: calculating the inverse

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

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

$$= \frac{1}{(-5)} \begin{bmatrix} -1 & -3 \\ -1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1/5 & 3/5 \\ 1/5 & -2/5 \end{bmatrix}$$

Sanity check: $\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} 1/5 & 3/5 \\ 1/5 & -2/5 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} + \frac{3}{5} & \frac{3}{5} - \frac{3}{5} \\ \frac{2}{5} - \frac{2}{5} & \frac{3}{5} + \frac{2}{5} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Note: calculating the inverse is usually done by computer programs

A unique inverse cannot be calculated if the determinate is zero,
[1] the matrix is said to be singular
[2] need to use a generalized inverse

Matrix operations: example

We have measured the weight of 6 people, what is their mean weight?

Let $\boldsymbol{w}$ be the 1x6 row vector of weights

$\mathbf{1}'\boldsymbol{w}$ gives the sum of the elements of $\boldsymbol{w}$

$\mathbf{1}'\boldsymbol{w}/\boldsymbol{n}$ is the average of the elements of $\boldsymbol{w}$

$$\boldsymbol{w} = [53 \quad 60 \quad 85 \quad 70 \quad 72 \quad 64]$$

$$\mathbf{1}'\boldsymbol{w} = 404$$

$$\mathbf{1}'\boldsymbol{w}/\mathbf{6} = 67.3$$

Matrix operations: example

e.g. simultaneous equations

2x + 3y = 15

x - y = 1

can be rewritten as $\mathbf{Av} = \mathbf{b}$; where $\mathbf{A} = \begin{bmatrix} \mathbf{2} & \mathbf{3} \\ \mathbf{1} & \mathbf{-1} \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ & $\mathbf{b} = \begin{bmatrix} \mathbf{15} \\ \mathbf{1} \end{bmatrix}$

What values are x and y?

$$\mathbf{Av} = \mathbf{b}$$
$$A^{-1}\mathbf{Av} = A^{-1}\mathbf{b}$$
$$\mathbf{v} = A^{-1}\mathbf{b}$$

$$= \begin{bmatrix} \mathbf{0.2} & \mathbf{0.6} \\ \mathbf{0.2} & \mathbf{-0.4} \end{bmatrix} \begin{bmatrix} \mathbf{15} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{0.2}x\mathbf{15} + \mathbf{0.6}x\mathbf{1} \\ \mathbf{0.2}x\mathbf{15} - \mathbf{0.4}x\mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{3.6} \\ \mathbf{2.6} \end{bmatrix}$$

PCA analysis

- PCA (Principal component analysis) is a useful method to summarize the ‘essence’ of your data
  - ➢Dimension reduction technique
- Based on *eigen-decomposition* of a diagonalizable matrix
- Exact methods (R/GCTA/EIGENSTRAT) are computationally expensive for large datasets
  - Approximate methods, e.g. fastPCA

- Good 20min step-by-step intro PCA analysis on YouTube, StatQuest
  - ➢https://www.youtube.com/watch?v=FgakZw6K1QQ

Example: PCA analysis in UK Biobank

- Bycroft et al 2018 Nature
- Genomic relationship matrix is a n x n matrix of SNP relationships between individuals
- Software: fastPCA
  - also possible in GCTA for smaller datasets

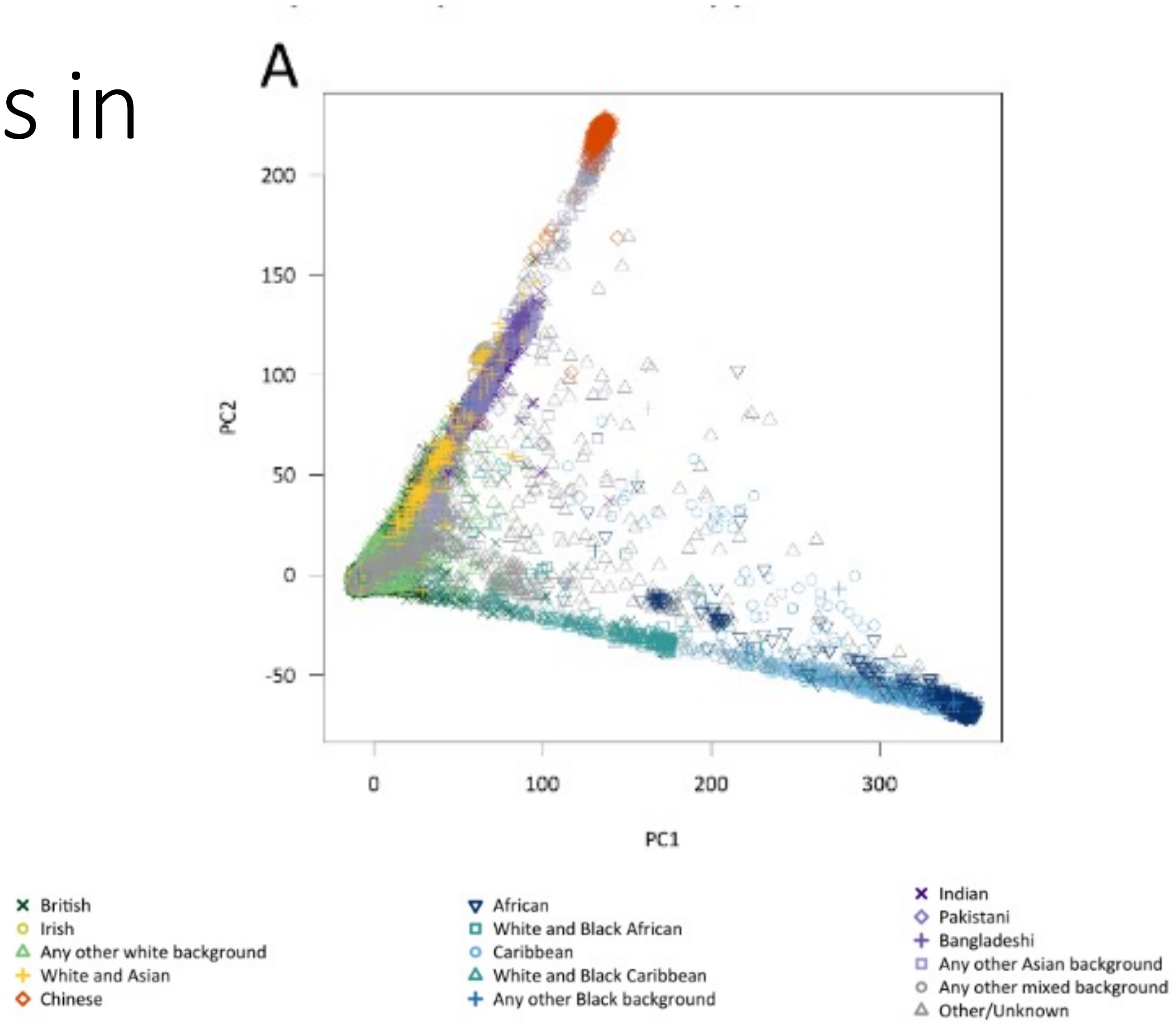


Figure 2 Genetic principal components in UK Biobank, computed from 141,0670 samples and 101,284 SNPs using flashPCA [10]. **(A)** The 1st principal component (PC1) on the *x*-axis and the 2nd principal component (PC2) on the *y*-axis. **(B)** The 3rd principal component (PC3) on the *x*-axis and the 4th principal component (PC4) on the *y*-axis. In both panels, samples are coloured according to self-reported ethnicity. The legend indicates the coloured symbol used for each predefined ethnicity throughout this document.

A brief introduction to: Eigenvalues and eigenvectors

- Given a square matrix **A**,
  - Can we define a vector **v** and scalar **b** such that **Av** = b**v**? If so then b is an eigenvalue and **v** an eigenvector of **A**.
  - Note:
    - Eigen values and eigen vectors work as a 'pair'
    - there maybe 0 to n (where n is the rank of **A**) eigenvalue/eigenvector pairs for **A**
    - e.g.

Is $\mathbf{v}$= $\begin{bmatrix} \mathbf{2} \\ \mathbf{1} \end{bmatrix}$ and b =4 an eigenvector/eigenvalue pair of $\mathbf{A}$ = $\begin{bmatrix} \mathbf{3} & \mathbf{2} \\ \mathbf{3} & \mathbf{-2} \end{bmatrix}$?

b**v = 4** $\begin{bmatrix} \mathbf{2} \\ \mathbf{1} \end{bmatrix}$= $\begin{bmatrix} \mathbf{8} \\ \mathbf{4} \end{bmatrix}$

$\mathbf{Av}$= $\begin{bmatrix} \mathbf{3} & \mathbf{2} \\ \mathbf{3} & \mathbf{-2} \end{bmatrix} \begin{bmatrix} \mathbf{2} \\ \mathbf{1} \end{bmatrix}$

= $\begin{bmatrix} \mathbf{3x2 + 2x1} \\ \mathbf{3x2 - 2x1} \end{bmatrix}$= $\begin{bmatrix} \mathbf{8} \\ \mathbf{4} \end{bmatrix}$

Eigenvalues, eigenvectors & PCA analysis

- In PCA,
  - Eigenvalues are ordered, largest to smallest
  - Eigenvectors are orthogonal
- PC1 (1^{st} principal component) has the largest eigenvalue. PC1 represents the largest axis of variation in the matrix.
- PC2 is orthogonal (uncorrelated) to PC1 and has the 2nd largest eigenvalue. It represents the 2^{nd} largest axis of variation in the matrix.
  - ...etc.
- e.g. $\mathbf{K}$ is decomposed into eigenvalues $\mathbf{D}$ and eigenvectors $\mathbf{E}$, then $\mathbf{K} = \mathbf{EDE}'$

Eigenvalues, eigenvectors & PCA analysis

What is the primary axis of variation in $\mathbf{K}$, if $\boldsymbol{K} = \begin{bmatrix} 1348 & 66.5 & -117.7 \\ 66.5 & 24.3 & -14.0 \\ -111.7 & -14.0 & 14.5 \end{bmatrix}$

Let $\mathbf{K} = \mathbf{EDE}'$

where **D** is the eigenvalues, $\boldsymbol{D} = \begin{bmatrix} 1361 & 0 & 0 \\ 0 & 24.5 & 0 \\ 0 & 0 & 1.5 \end{bmatrix}$

and **E** are the eigenvectors, $\boldsymbol{E} = \begin{bmatrix} -0.995 & -0.079 & 0.056 \\ -0.050 & 0.915 & 0.400 \\ 0.083 & -0.395 & 0.915 \end{bmatrix}$

basics – practical 1

- Some matrix algebra to do by-hand (!)

- Work through “basicsPrac1.pdf”
  - PART 1: check answers to matrix algebra questions
  - PART 2: Small example PCA analysis

- Software: R

Cluster Access

- You have all been provided with login details to computing resources needed for the practical component

- An SSH terminal is needed to connect to the computing:
    - Windows: Install PuTTY
    - Hostname: as provided (203.101.228.xxx)
    - User: as provided
    - Check Connection > SSH > X11 > Enable X11 forwarding

    - Mac/Linux: Use the terminal
    - `ssh -X <user>@203.101.228.xxx`