An Introduction to Matrix Algebra

EPSY 905: Fundamentals of Multivariate Modeling

Online Lecture #8

In This Lecture…

- An introduction to matrix algebra
  - Scalars, vectors, and matrices
  - Basic matrix operations
  - Advanced matrix operations

- An introduction to matrices in R
  - Embedded within the R language

Why Learning a Little Matrix Algebra is Important

- Matrix algebra is the alphabet of the language of statistics
  - You will most likely encounter formulae with matrices very quickly

- For example, imagine you were interested in analyzing some repeated measures data…but things don't go as

Formulation of the Mixed Model

The previous general linear model is certainly a useful one (Sea
although you still assume normality.

The mixed model is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

where everything is the same as in the general linear model ex
Henderson (1990) and Searle, Casella, and McCulloch (1992) f

A key assumption in the foregoing analysis is that $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are

$$\mathrm{E}\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$$

$$\mathrm{Var}\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$$

The variance of $\mathbf{y}$ is, therefore, $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$. You can model

Estimating Covariance Parameters in the Mixed Model

Estimation is more difficult in the mixed model than in the general linear model. Not only do y

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

However, it requires knowledge of $\mathbf{V}$ and, therefore, knowledge of $\mathbf{G}$ and $\mathbf{R}$. Lacking such info

In many situations, the best approach is to use *likelihood-based* methods, exploiting the ass
(REML). A favorable theoretical property of ML and REML is that they accommodate data tha

PROC MIXED constructs an objective function associated with ML or REML and maximizes i

$$\text{ML}: \quad l(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\mathbf{r}'\mathbf{V}^{-1}\mathbf{r} - \frac{n}{2}\log(2\pi)$$

$$\text{REML}: \quad l_R(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{r}'\mathbf{V}^{-1}\mathbf{r} - \frac{n-p}{2}\log(2\pi)\}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and p is the rank of $\mathbf{X}$. PROC MIXED actually minimiz
analytical details for implementing a QR-decomposition approach to the problem. Wolfinger, T

Introduction and Motivation

- Nearly all multivariate statistical techniques are described with matrix algebra

- When new methods are developed, the first published work typically involves matrices
  - It makes technical writing more concise – formulae are smaller

- Have you seen:
  - $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$
  - $\boldsymbol{\Lambda\Phi\Lambda}^T + \boldsymbol{\Psi}$

- **Useful tip:** matrix algebra is a great way to get out of boring conversations and other awkward moments

Definitions

- We begin this class with some general definitions (from dictionary.com):
  - **Matrix:**
    1. A rectangular array of numeric or algebraic quantities subject to mathematical operations
    2. The substrate on or within which a fungus grows

  - **Algebra:**
    1. A branch of mathematics in which symbols, usually **letters** of the alphabet, represent numbers or members of a specified set and are used to represent quantities and to express general relationships that hold for all members of the set

    2. A set together with a pair of **binary operations** defined on the set. Usually, the set and the operations include an **identity element**, and the operations are **commutative** or **associative**

Why Learn Matrix Algebra

- Matrix algebra can seem very abstract from the purposes of this class (and statistics in general)

- Learning matrix algebra is important for:
  - Understanding how statistical methods work
    - And when to use them (or not use them)
  - Understanding what statistical methods mean
  - Reading and writing results from new statistical methods

- This is a first lecture of learning the language of multivariate statistics

DATA EXAMPLE AND R

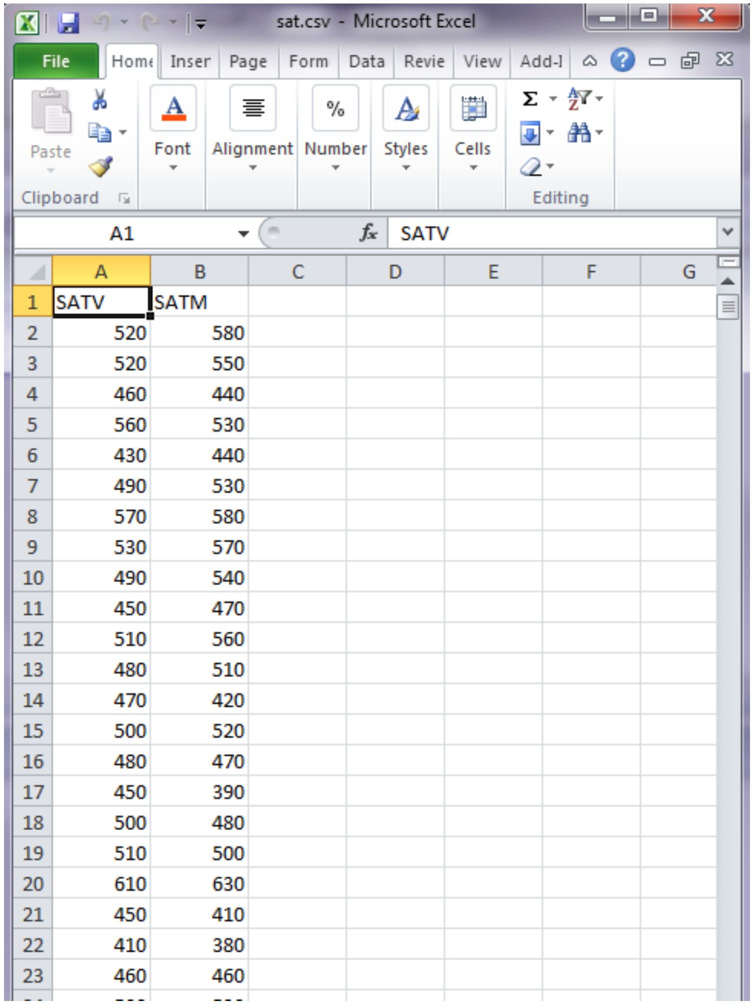
A Guiding Example

- To demonstrate matrix algebra, we will make use of data
- Imagine that I collected data SAT test scores for both the Math (SATM) and Verbal (SATV) sections of 1,000 students
- The descriptive statistics of this data set are given below:

Statistic	SATV	SATM
Mean	499.3	498.3
SD	49.8	81.2
Correlation		
SATV	1.00	0.78
SATM	0.78	1.00

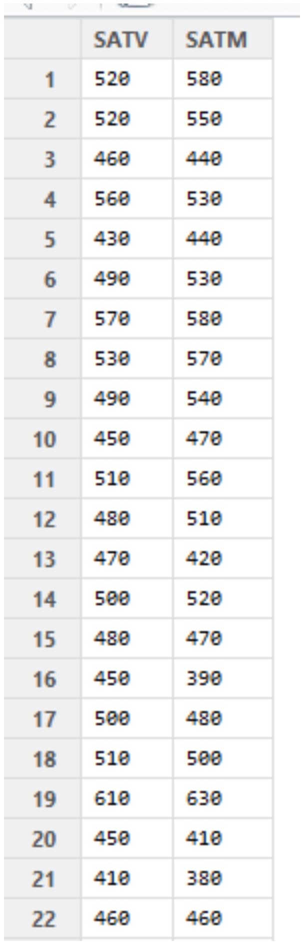
The Data…

In Excel:



	A	B
1	SATV	SATM
2	520	580
3	520	550
4	460	440
5	560	530
6	430	440
7	490	530
8	570	580
9	530	570
10	490	540
11	450	470
12	510	560
13	480	510
14	470	420
15	500	520
16	480	470
17	450	390
18	500	480
19	510	500
20	610	630
21	450	410
22	410	380
23	460	460

In R:



	SATV	SATM
1	520	580
2	520	550
3	460	440
4	560	530
5	430	440
6	490	530
7	570	580
8	530	570
9	490	540
10	450	470
11	510	560
12	480	510
13	470	420
14	500	520
15	480	470
16	450	390
17	500	480
18	510	500
19	610	630
20	450	410
21	410	380
22	460	460

DEFINITIONS OF MATRICES, VECTORS, AND SCALARS

Matrices

- A matrix is a rectangular array of data
  - Used for storing numbers
- Matrices can have unlimited dimensions
  - For our purposes all matrices will have two dimensions:
    - Row
    - Columns
- Matrices are symbolized by **boldface** font in text, typically with capital letters
  - Size (r rows x c columns)

SAT Verbal (Column 1) SAT Math (Column 2)

$$\mathbf{X} = \begin{bmatrix} 520 & 580 \\ 520 & 550 \\ \vdots & \vdots \\ 540 & 660 \end{bmatrix}_{(1000\ x\ 2)}$$

Vectors

- A vector is a matrix where one dimension is equal to size 1
  - Column vector: a matrix of size $r \ x \ 1$

$$\boldsymbol{x}_{\cdot 1} = \begin{bmatrix} 520 \\ 520 \\ \vdots \\ 540 \end{bmatrix}_{1000 \ x \ 1}$$

  - Row vector: a matrix of size $1 \ x \ c$

$$\boldsymbol{x}_{1\cdot} = \begin{bmatrix} 520 & 580 \end{bmatrix}_{1 \ x \ 2}$$

- Vectors are typically written in **boldface** font text, usually with lowercase letters

- The dots in the subscripts $\boldsymbol{x}_{\cdot 1}$ and $\boldsymbol{x}_{1\cdot}$ represent the dimension aggregated across in the vector
  - $\boldsymbol{x}_{1\cdot}$ is the first row and **all** columns of $\mathbf{X}$
  - $\boldsymbol{x}_{\cdot 1}$ is the first column and **all** rows of $\mathbf{X}$
  - Sometimes the rows and columns are separated by a comma (making it possible to read double-digits in either dimension)

Matrix Elements

- A matrix (or vector) is composed of a set of elements
  - Each element is denoted by its position in the matrix (row and column)

- For our matrix of data $\mathbf{X}$ (size 1000 rows and 2 columns), each element is denoted by:

$$x_{ij}$$

  - The first subscript is the index for the rows: i = 1,…,r (= 1000)
  - The second subscript is the index for the columns: j = 1,…,c (= 2)

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{1000,1} & x_{1000,2} \end{bmatrix}_{(1000\ x\ 2)}$$

Scalars

- A scalar is just a single number

- The name scalar is important: the number "scales" a vector – it can make a vector "longer" or "shorter"

- Scalars are typically written without boldface:

$$x_{11} = 520$$

- Each element of a matrix is a scalar

Matrix Transpose

- The transpose of a matrix is a reorganization of the matrix by switching the indices for the rows and columns

$$\mathbf{X} = \begin{bmatrix} 520 & 580 \\ 520 & 550 \\ \vdots & \vdots \\ 540 & 660 \end{bmatrix}_{(1000 \; x \; 2)}$$

$$\mathbf{X}^T = \begin{bmatrix} 520 & 520 & \cdots & 540 \\ 580 & 550 & \cdots & 660 \end{bmatrix}_{(2 \; x \; 1000)}$$

- An element x_{ij} in the original matrix $\mathbf{X}$ is now x_{ji} in the transposed matrix $\mathbf{X}^T$

- **Transposes are used to align matrices for operations where the sizes of matrices matter (such as matrix multiplication)**

Types of Matrices

- **Square Matrix:** A square matrix has the same number of rows and columns
  - Correlation/covariance matrices are square matrices

- **Diagonal Matrix:** A diagonal matrix is a square matrix with non-zero diagonal elements ($x_{ij} \neq 0 \text{ for } i = j$) and zeros on the off-diagonal elements ($x_{ij} = 0 \text{ for } i \neq j$):

$$\mathbf{A} = \begin{bmatrix} 2.759 & 0 & 0 \\ 0 & 1.643 & 0 \\ 0 & 0 & 0.879 \end{bmatrix}$$

  - We will use diagonal matrices to form correlation matrices

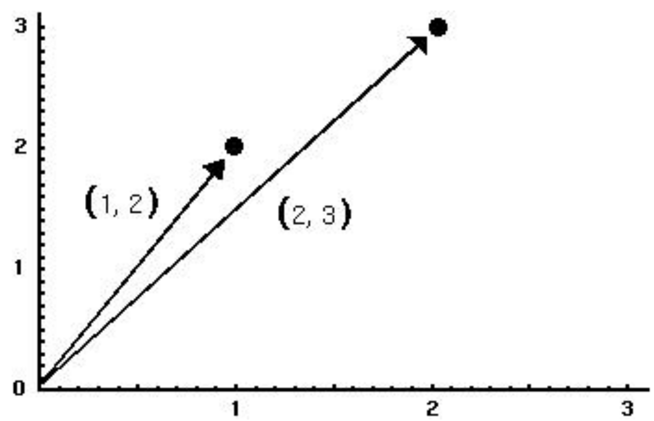
- **Symmetric Matrix:** A symmetric matrix is a square matrix where all elements are reflected across the diagonal ($a_{ij} = a_{ji}$)
  - Correlation and covariance matrices are symmetric matrices

THE UNIVERSITY OF KANSAS

VECTORS

Vectors in Space…

- Vectors (row or column) can be represented as lines on a Cartesian coordinate system (a graph)
- Consider the vectors: $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$
- A graph of these vectors would be:



- **<u>Question:</u>** how would a column vector for each of our example variables (SATM and SATV) be plotted?

Vector Length

- The length of a vector emanating from the origin is given by the Pythagorean formula
  - This is also called the Euclidean distance between the endpoint of the vector and the origin

$$L_{\mathbf{x}} = \sqrt{x_{11}^2 + x_{21}^2 + \cdots + x_{r1}^2} = ||\mathbf{x}||$$

- From the last slide: $||\mathbf{a}|| = \sqrt{5} = 2.24; ||\mathbf{b}|| = \sqrt{13} = 3.61$

- From our data:

$$||\mathbf{SATV}|| = 15{,}868.138; ||\mathbf{SATM}|| = 15{,}964.42$$

- **<u>In data:</u>** length is an analog to the standard deviation
  - In mean-centered variables, the length is the square root of the sum of mean deviations (not quite the SD, but close)

Vector Addition

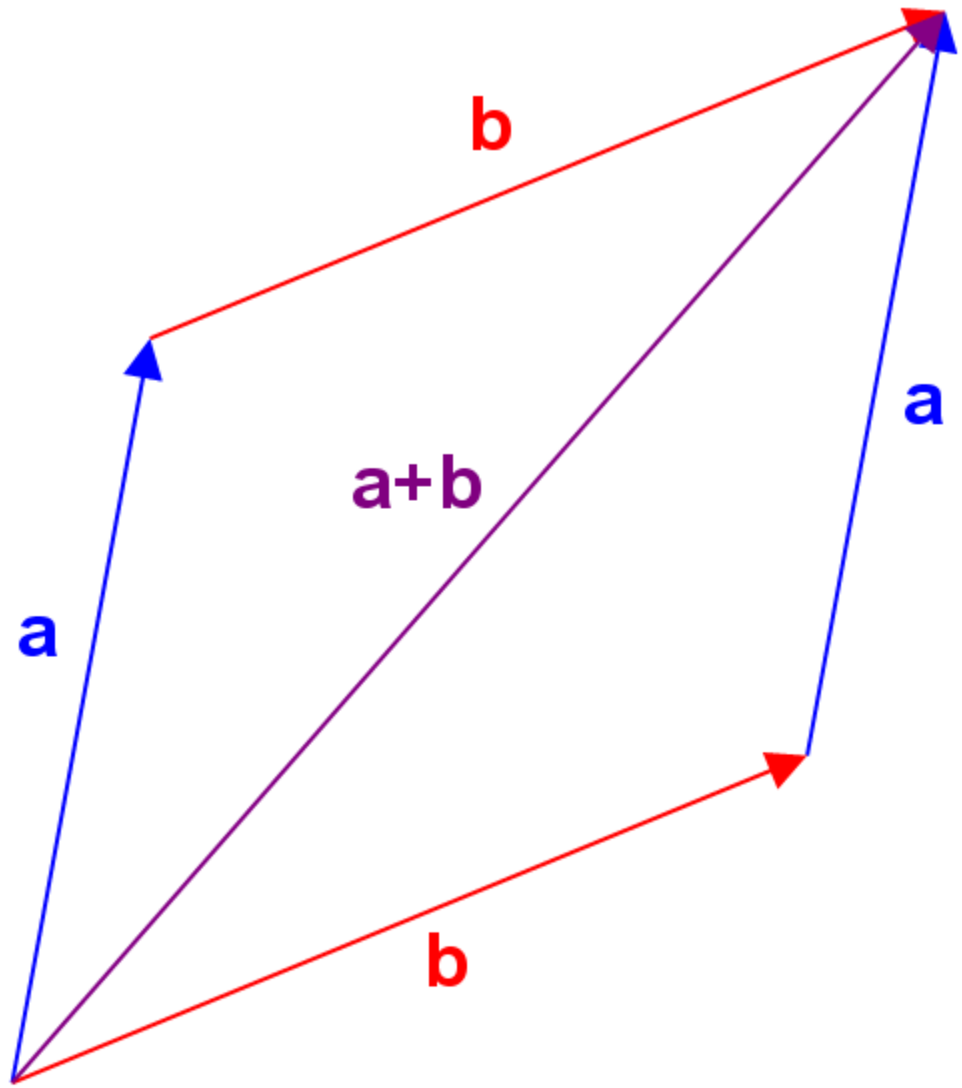
- Vectors can be added together so that a new vector is formed
- Vector addition is done element-wise, by adding each of the respective elements together:
  - The new vector has the same number of rows and columns

$$\mathbf{c} = \mathbf{a} + \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

  - Geometrically, this creates a new vector along either of the previous two
    - Starting at the origin and ending at a new point in space
- <u>**In Data:**</u> a new variable (say, SAT total) is the result of vector addition

$$\boldsymbol{SAT_{TOTAL}} = \boldsymbol{x}_{.1} + \boldsymbol{x}_{.2}$$

Vector Addition: Geometrically



Vector Multiplication by Scalar

- Vectors can be multiplied by scalars
  - All elements are multiplied by the scalar

$$\mathbf{d} = 2\mathbf{a} = 2\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}2\\4\end{bmatrix}$$

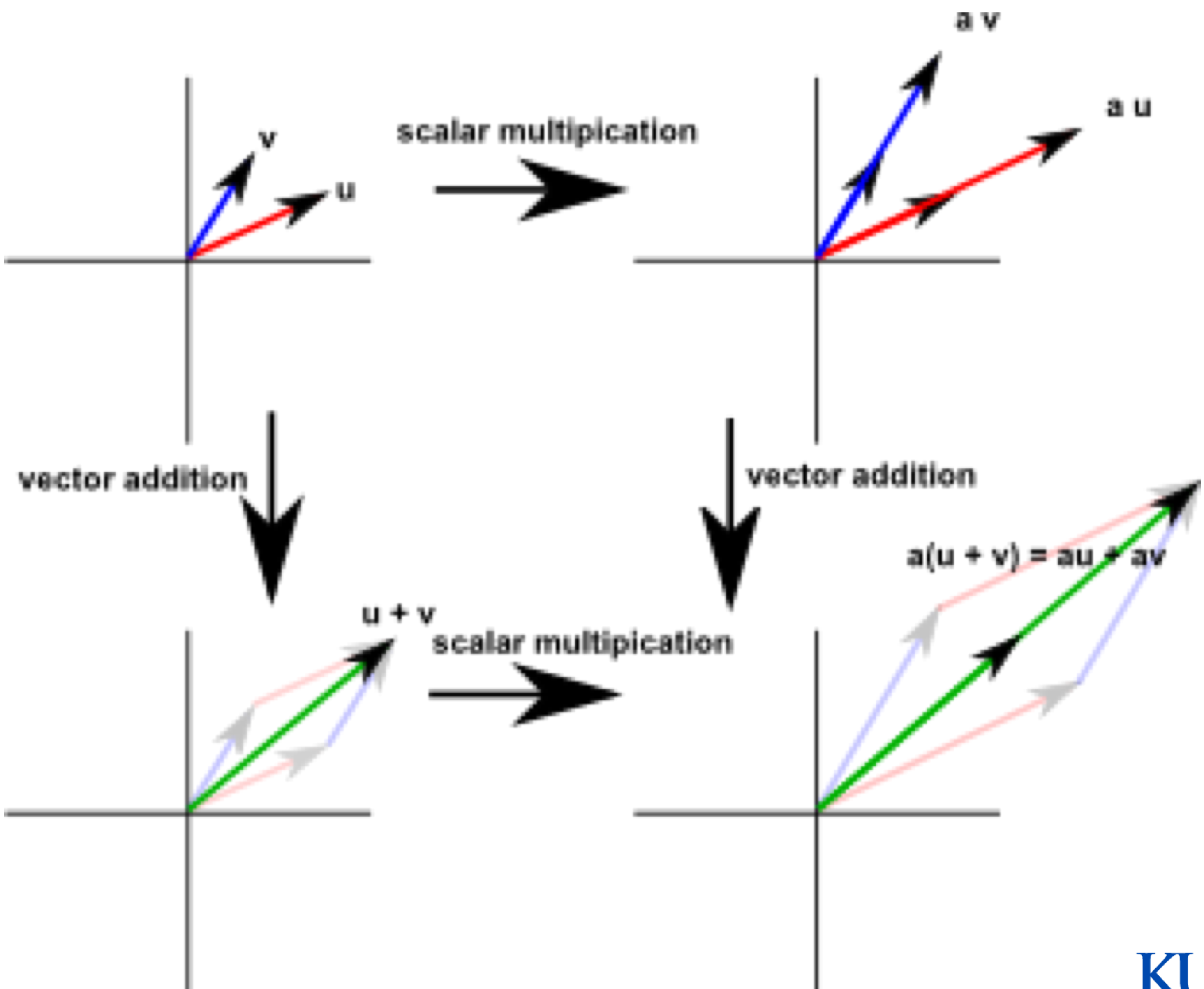
- Scalar multiplication changes the length of the vector:

$$||\mathbf{d}|| = \sqrt{2^2 + 4^2} = \sqrt{20} = 4.47$$

- This is where the term scalar comes from: a scalar ends up "rescaling" (resizing) a vector

- **<u>In Data:</u>** the GLM (where **X** is a matrix of data) the fixed effects (slopes) are scalars multiplying the data

Scalar Multiplication: Geometrically



Linear Combinations

- Addition of a set of vectors (all multiplied by scalars) is called a linear combination:

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_k\mathbf{x}_k$$

- Here, $\mathbf{y}$ is the linear combination
  - For all *k* vectors, the set of all possible linear combinations is called their span
  - Typically not thought of in most analyses – but when working with things that don't exist (latent variables) becomes somewhat important

- **<u>In Data</u>**: linear combinations happen frequently:
  - Linear models (i.e., Regression and ANOVA)
  - Principal components analysis

Linear Dependencies

- A set of vectors are said to be linearly dependent if

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_k\mathbf{x}_k = 0$$

-and-

$a_1, a_2, \ldots, a_k$ are all **not** zero

- Example: let's make a new variable – SAT Total:

$$\mathbf{SAT_{total}} = 1 * \mathbf{SATV} + 1 * \mathbf{SATM}$$

- The new variable is linearly dependent with the others:

$$(1) * \mathbf{SATV} + (1) * \mathbf{SATM} + (-1) * \mathbf{SAT_{total}} = \mathbf{0}$$

- **In Data**: (multi)collinearity is a linear dependency. Linear dependencies are bad for statistical analyses that use matrix inverses

Inner (Dot) Product of Vectors

- An important concept in vector geometry is that of the inner product of two vectors
  - The inner product is also called the dot product

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T\mathbf{b} = a_{11}b_{11} + a_{21}b_{21} + \cdots + a_{N1}b_{N1} = \sum_{i=1}^{N} a_{i1}b_{i1}$$

- The dot or inner product is related to the angle between vectors and to the projection of one vector onto another

- From our example: $\mathbf{a} \cdot \mathbf{b} = 1 * 2 + 2 * 3 = 8$
- From our data: $\boldsymbol{x}_{.1} \cdot \boldsymbol{x}_{.2} = 251{,}928{,}400$

- **<u>In data:</u>** the angle between vectors is related to the correlation between variables and the projection is related to regression/ANOVA/linear models

Angle Between Vectors

- As vectors are conceptualized geometrically, the angle between two vectors can be calculated

$$\theta_{ab} = \cos^{-1}\left(\frac{\mathbf{a} \cdot \boldsymbol{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}\right)$$

- From the example:

$$\theta_{ab} = \cos^{-1}\left(\frac{8}{\sqrt{5}\sqrt{13}}\right) = 0.12$$

- From our data:

$$\theta_{SATV,SATM} = \cos^{-1}\left(\frac{251{,}928{,}400}{\sqrt{15{,}868.138}\sqrt{15{,}946.42}}\right) = 0.105$$

In Data: Cosine Angle = Correlation

- If you have data that are:
  - Placed into vectors
  - Centered by the mean (subtract the mean from each observation)

- ...then the cosine of the angle between those vectors is the correlation between the variables:

$$r_{ab} = cos(\theta_{ab}) = \frac{\mathbf{a} \cdot \boldsymbol{b}}{\|\mathbf{a}\|\|\mathbf{b}\|} = \frac{\sum_{i=1}^{N}(a_{i1} - \bar{a})\left(b_{i1} - \bar{b}\right)}{\sqrt{\sum_{i=1}^{N}(a_{i1} - \bar{a})^2}\sqrt{\sum_{i=1}^{N}\left(b_{i1} - \bar{b}\right)^2}}$$

For the SAT example data (using mean centered variables):

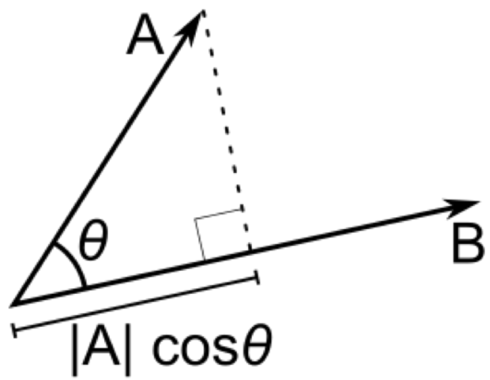
$$r_{SATV,SATM} = \cos\left(\theta_{SATVc,SATMc}\right)$$
$$= \cos\left(\frac{3{,}132{,}223.6}{1{,}573.956 * 2{,}567.0425}\right) = .775$$

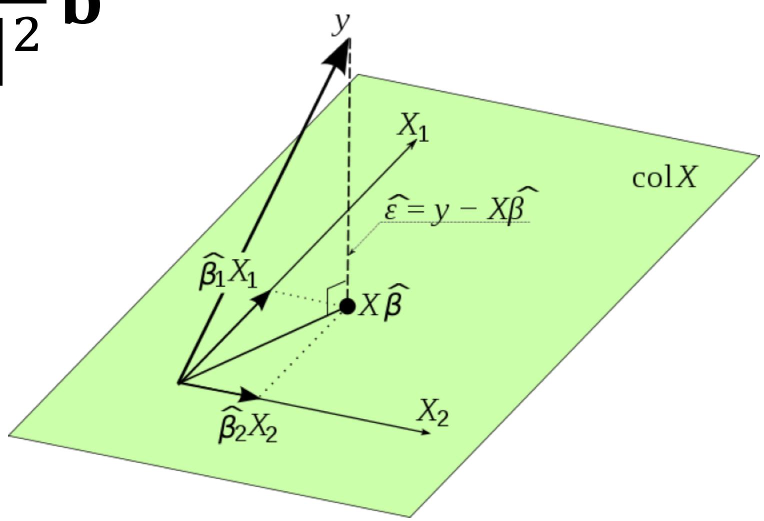
Vector Projections

- A final vector property that shows up in statistical terms frequently is that of a projection

- The **projection** of a vector $\mathbf{a}$ onto $\mathbf{b}$ is the orthogonal projection of $\mathbf{a}$ onto the straight line defined by $\mathbf{b}$
  - The projection is the "shadow" of one vector onto the other:

$$\mathbf{a}_{\text{proj } \mathbf{b}} = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2}\mathbf{b}$$





- **<u>In data:</u>** linear models can be thought of as projections

Vector Projections Example

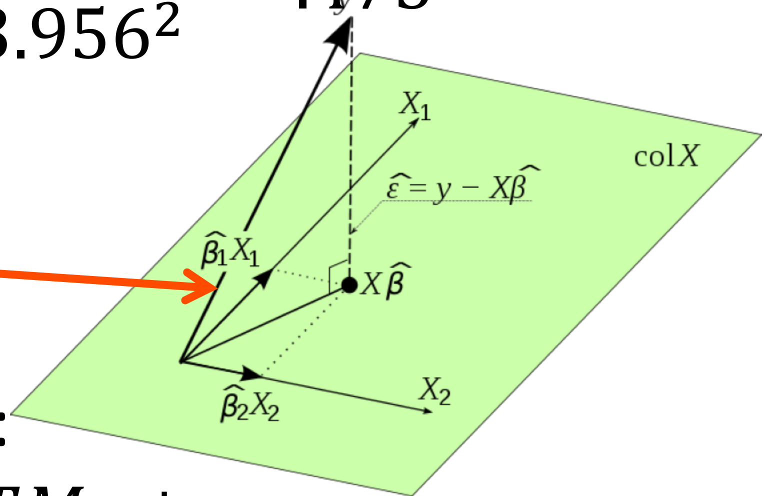
- To provide a bit more context for vector projections, let's consider the projection of mean centered SATV onto SATM:

$$\boldsymbol{SATVc}_{\text{proj}\,\boldsymbol{SATMc}} = \frac{\boldsymbol{SATVc} \cdot \boldsymbol{SATMc}}{\|\boldsymbol{SATMc}\|^2}\boldsymbol{SATMc}$$

- The first portion turns out to be:

$$\frac{\boldsymbol{SATVc} \cdot \boldsymbol{SATMc}}{\|\boldsymbol{SATMc}\|^2} = \frac{3{,}132{,}223.6}{1{,}573.956^2} = .475$$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 262.48200    6.18941   42.41   <2e-16 ***
satm          0.47532    0.01226   38.77   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



- This is also the regression slope β_1:

$$SATV_p = \beta_0 + \beta_1 SATM_p + e_p$$

THE UNIVERSITY OF KANSAS

MATRIX ALGEBRA

Moving from Vectors to Matrices

- A matrix can be thought of as a collection of vectors
  - Matrix operations are vector operations on steroids
- Matrix algebra defines a set of operations and entities on matrices
  - I will present a version meant to mirror your previous algebra experiences
- Definitions:
  - Identity matrix
  - Zero vector
  - Ones vector
- Basic Operations:
  - Addition
  - Subtraction
  - Multiplication
  - “Division”

Matrix Addition and Subtraction

- Matrix addition and subtraction are much like vector addition/subtraction

- Rules:
  - Matrices must be the same size (rows and columns)

- Method:
  - The new matrix is constructed of element-by-element addition/subtraction of the previous matrices

- Order:
  - The order of the matrices (pre- and post-) does not matter

Matrix Addition/Subtraction

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \\ a_{31} + b_{31} & a_{32} + b_{32} \\ a_{41} + b_{41} & a_{42} + b_{42} \end{bmatrix} \qquad \mathbf{A} - \mathbf{B} = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} \\ a_{21} - b_{21} & a_{22} - b_{22} \\ a_{31} - b_{31} & a_{32} - b_{32} \\ a_{41} - b_{41} & a_{42} - b_{42} \end{bmatrix}$$

Matrix Multiplication

- Matrix multiplication is a bit more complicated
  - The new matrix may be a different size from either of the two multiplying matrices

$$\mathbf{A}_{(r\ x\ c)}\mathbf{B}_{(c\ x\ k)} = \mathbf{C}_{(r\ x\ k)}$$

- Rules:
  - Pre-multiplying matrix must have number of columns equal to the number of rows of the post-multiplying matrix
- Method:
  - The elements of the new matrix consist of the inner (dot) product of the row vectors of the pre-multiplying matrix and the column vectors of the post-multiplying matrix
- Order:
  - The order of the matrices (pre- and post-) matters

Matrix Multiplication

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} & a_{31}b_{13} + a_{32}b_{23} \\ a_{41}b_{11} + a_{42}b_{21} & a_{41}b_{12} + a_{42}b_{22} & a_{41}b_{13} + a_{42}b_{23} \end{bmatrix}$$

Multiplication in Statistics

- Many statistical formulae with summation can be re-expressed with matrices

- A common matrix multiplication form is: $\mathbf{X}^T\mathbf{X}$
  - Diagonal elements: $\sum_{i=1}^{N} X_i^2$
  - Off-diagonal elements: $\sum_{i=1}^{N} X_{ia}X_{ib}$

- For our SAT example:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} \sum_{i=1}^{N} SATV_i^2 & \sum_{i=1}^{N} SATV_i SATM_i \\ \sum_{i=1}^{N} SATV_i SATM_i & \sum_{i=1}^{N} SATM_i^2 \end{bmatrix}$$
$$= \begin{bmatrix} 251{,}797{,}800 & 251{,}928{,}400 \\ 251{,}928{,}400 & 254{,}862{,}700 \end{bmatrix}$$

THE UNIVERSITY OF KANSAS

Identity Matrix

- The identity matrix is a matrix that, when pre- or post-multiplied by another matrix results in the original matrix:

$$\mathbf{AI} = \mathbf{A}$$
$$\mathbf{IA} = \mathbf{A}$$

- The identity matrix is a square matrix that has:
  - Diagonal elements = 1
  - Off-diagonal elements = 0

$$I_{(3\ x\ 3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Zero Vector

- The zero vector is a column vector of zeros

$$\mathbf{0}_{(3\ x\ 1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- When pre- or post- multiplied the result is the zero vector:

$$\mathbf{A0} = \mathbf{0}$$
$$\mathbf{0A} = \mathbf{0}$$

Ones Vector

- A ones vector is a column vector of 1s:

$$\mathbf{1}_{(3\ x\ 1)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- The ones vector is useful for calculating statistical terms, such as the mean vector and the covariance matrix

Matrix “Division”: The Inverse Matrix

- Division from algebra:
  - First: $\frac{a}{b} = \frac{1}{b}a = b^{-1}a$
  - Second: $\frac{a}{a} = 1$
- “Division” in matrices serves a similar role
  - For square and symmetric matrices, an inverse matrix is a matrix that when pre- or post- multiplied with another matrix produces the identity matrix:

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

- Calculation of the matrix inverse is complicated
  - Even computers have a tough time
- Not all matrices can be inverted
  - Non-invertible matrices are called singular matrices
    - In statistics, singular matrices are commonly caused by linear dependencies

The Inverse

- **In data:** the inverse shows up constantly in statistics
  - Models which assume some type of (multivariate) normality need an inverse covariance matrix

- Using our SAT example
  - Our data matrix was size (1000 x 2), which is not invertible
  - However $\mathbf{X}^T\mathbf{X}$ was size (2 x 2) – square, and symmetric

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 251{,}797{,}800 & 251{,}928{,}400 \\ 251{,}928{,}400 & 254{,}862{,}700 \end{bmatrix}$$

  - The inverse is:

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 3.61E-7 & -3.57E-7 \\ -3.57E-7 & 3.56E-7 \end{bmatrix}$$

Matrix Algebra Operations

- $(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$
- $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$
- $c(\mathbf{A}+\mathbf{B})=c\mathbf{A}+c\mathbf{B}$
- $(c+d)\mathbf{A}=c\mathbf{A}+d\mathbf{A}$
- $(\mathbf{A}+\mathbf{B})^T=\mathbf{A}^T+\mathbf{B}^T$
- $(cd)\mathbf{A}=c(d\mathbf{A})$
- $(c\mathbf{A})^T=c\mathbf{A}^T$
- $c(\mathbf{AB})=(c\mathbf{A})\mathbf{B}$
- $\mathbf{A}(\mathbf{BC})=(\mathbf{AB})\mathbf{C}$
- $\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$
- $(\mathbf{AB})^T=\mathbf{B}^T\mathbf{A}^T$
- For x_j such that Ax_j exists:

$$\sum_{j=1}^{N}\mathbf{A}\mathbf{x}_j=\mathbf{A}\sum_{j=1}^{N}\mathbf{x}_j$$

$$\sum_{j=1}^{N}(\mathbf{A}\mathbf{x}_j)(\mathbf{A}\mathbf{x}_j)^T=\mathbf{A}\left(\sum_{j=1}^{N}\mathbf{x}_j\mathbf{x}_j^T\right)\mathbf{A}^T$$

ADVANCED MATRIX OPERATIONS

Advanced Matrix Functions/Operations

- We end our matrix discussion with some advanced topics
  - All related to multivariate statistical analysis

- To help us throughout, let's consider the correlation matrix of our SAT data:

$$\mathbf{R} = \begin{bmatrix} 1.00 & 0.78 \\ 0.78 & 1.00 \end{bmatrix}$$

Matrix Trace

- For a square matrix $\mathbf{A}$ with p rows/columns, the trace is the sum of the diagonal elements:

$$tr\mathbf{A} = \sum_{i=1}^{p} a_{ii}$$

- For our data, the trace of the correlation matrix is 2
  - For all correlation matrices, the trace is equal to the number of variables because all diagonal elements are 1

- The trace is considered the total variance in multivariate statistics
  - Used as a target to recover when applying statistical models

Matrix Determinants

- A square matrix can be characterized by a scalar value called a determinant:

$$\det \mathbf{A} = |\mathbf{A}|$$

- Calculation of the determinant is tedious
  - Our determinant was 0.3916
- The determinant is useful in statistics:
  - Shows up in multivariate statistical distributions
  - Is a measure of “generalized” variance of multiple variables
- If the determinant is positive, the matrix is called **positive definite → the matrix has an inverse**
- If the determinant is not positive, the matrix is called **non-positive definite → the matrix does not have an inverse**

WRAPPING UP

Much Ado About Matrices…

- Matrices show up nearly anytime multivariate statistics are used, often in the help/manual pages of the package you intend to use for analysis
- You don’t have to do matrix algebra, but please do try to understand the concepts underlying matrices
- Your working with multivariate statistics will be better off because of even a small amount of understanding