Matrix Algebra Review & GLM: General Linear Models

Advanced Biostatistics
Dean Adams
Lecture 5
EEOB 590C

Grasping matrix algebra is the ***KEY*** for understanding statistics

Compact method of expressing mathematical operations

Generalize from one to many variables (i.e. vectors to matrices)

Matrix operations have geometric interpretations in data spaces

Much of our data (e.g., shape) cannot be measured with a single variable, so multivariate methods are required to properly address our hypotheses (e.g., can evaluate covariance)

Scalar: a number

Vector: an ordered list (array) of scalars (n_{rows} x 1_{cols})

Column vectors are default

Matrix: a rectangular array of scalars (n_{rows} x p_{cols})

$$a = (3) \qquad \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix}$$

Reverse rows and columns

Represent by A^t or A'

Vector transpose works identically

$$\mathbf{A} = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$$

$$\mathbf{A}^{\mathbf{t}} = \mathbf{A}^{'} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

Matrices must have same dimensions

Add/subtract element-wise

Vector addition/subtraction works identically

Addition

$$\mathbf{A}+\mathbf{B} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} + \begin{bmatrix} 6 & 8 \\ 5 & 9 \end{bmatrix} = \begin{bmatrix} \mathbf{8} & \mathbf{12} \\ \mathbf{6} & \mathbf{12} \end{bmatrix}$$

Subtraction

$$\mathbf{A}-\mathbf{B} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} 6 & 8 \\ 5 & 9 \end{bmatrix} = \begin{bmatrix} \mathbf{-4} & \mathbf{-4} \\ \mathbf{-4} & \mathbf{-6} \end{bmatrix}$$

When multiplying a vector/matrix by a scalar, simply multiply all elements by the scalar

$$a\mathbf{e} = 3\begin{bmatrix} 2 \\ 7 \\ 5 \end{bmatrix} = \begin{bmatrix} 6 \\ 21 \\ 15 \end{bmatrix}$$

$$a\mathbf{B} = 5\begin{bmatrix} 1 & 2 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 0 & 25 \end{bmatrix}$$

Multiply 2 vectors:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Right?

$$\mathbf{ab} = \begin{bmatrix} a_1 b_1 \\ a_2 b_2 \\ \vdots \\ a_n b_n \end{bmatrix}$$

Wrong!

*Actually, not always wrong, but not general. The above is known as the 'Hadamard' product (element-wise multiplication), which is a special multiplication procedure that is only possible for vectors/matrices of the same dimension (and only done when specified as Hadamard product).

Combines multiplication and addition, and is performed in specified order

$$\mathbf{a}^{\mathrm{t}}\mathbf{b} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

Outer dimensions define resultant matrix

$$\overbrace{(1 \times \underbrace{3)(3}_{} \times 1)}^{}$$

Inner dimensions must match

Why inner product? In order for vector multiplication to work, dimensions have to match in a special way.

Inner product(or dot product) is a product that returns a scalar.

$$\mathbf{a}^{\mathrm{t}}\mathbf{b} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

Inner Product

$$\mathbf{b}\mathbf{a}^{\mathrm{t}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} a_1 b_1 & a_2 b_1 & \cdots & a_n b_1 \\ a_1 b_2 & a_2 b_2 & & \\ \vdots & & \ddots & \\ a_1 b_n & & & a_n b_n \end{bmatrix}$$

Outer Product

This is a “square matrix”
rows = columns

*Note the trace of $\mathbf{ba}^{\mathrm{t}}$ (the sum of diagonal elements) is equal to the inner product of $\mathbf{a}^{\mathrm{t}}\mathbf{b}$

Simply the joint multiplication of vectors within matrices

$$\mathbf{AB} = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 1 & -1 \\ & \end{bmatrix}\begin{bmatrix} 1 & \\ 1 & \end{bmatrix} = \begin{bmatrix} 1\cdot 1 - 1\cdot 1 & \\ & \end{bmatrix} = \begin{bmatrix} 0 & \\ & \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 1 & -1 \\ & \end{bmatrix}\begin{bmatrix} & 4 \\ & 1 \end{bmatrix} = \begin{bmatrix} & 1\cdot 4 - 1\cdot 1 \\ & \end{bmatrix} = \begin{bmatrix} & 3 \\ & \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} & \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 & \\ 1 & \end{bmatrix} = \begin{bmatrix} & \\ 2\cdot 1 + 3\cdot 1 & \end{bmatrix} = \begin{bmatrix} & \\ 5 & \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} & \\ 2 & 3 \end{bmatrix}\begin{bmatrix} & 4 \\ & 1 \end{bmatrix} = \begin{bmatrix} & \\ & 2\cdot 4 + 3\cdot 1 \end{bmatrix} = \begin{bmatrix} & \\ & 11 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ 5 & 11 \end{bmatrix}$$

Matrix multiplication is solving a series of inner products

Order of matrices makes a difference: **AB ≠BA**

$$\mathbf{AB} = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ 5 & 11 \end{bmatrix} \qquad \mathbf{BA} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 9 & 1 \\ 3 & 2 \end{bmatrix}$$

Also, matrix product will not always work in both directions:

$$\mathbf{AB} = \begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 10 & -4 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ 5 & 11 \\ 6 & 36 \end{bmatrix} \qquad \mathbf{BA} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 10 & -4 \end{bmatrix} = NA$$

(3 x 2)(2 x 2)
= (3 x 2)
inner numbers must match
then outer is the dimension

(2 x 2) (3 x 2)
= NOTHING
inner numbers don't match

Scalar multiplication:

$$a\mathbf{e} = 3\begin{bmatrix} 2 \\ 7 \\ 5 \end{bmatrix} = \begin{bmatrix} 6 \\ 21 \\ 15 \end{bmatrix}$$

Matrix multiplication:

$$\mathbf{AB} = \begin{bmatrix} \sum_{i=1}^{n} a_{1i} b_{i1} & \sum_{i=1}^{n} a_{1i} b_{i2} \\ \sum_{i=1}^{n} a_{2i} b_{i1} & \sum_{i=1}^{n} a_{2i} b_{i2} \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & 3 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} 2+6+12 & 1+6+6 \\ 8+15+24 & 4+15+12 \end{bmatrix} = \begin{bmatrix} \mathbf{20} & \mathbf{13} \\ \mathbf{47} & \mathbf{31} \end{bmatrix}$$

Inner dimensions MUST AGREE!!!

Matrix and description	Example	Comment
Square matrix Any matrix that has the same number of rows and columns (e.g., $n \times n$)	$\mathbf{S} = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 7 & 5 \\ 1 & 10 & 0 \end{bmatrix}$	These matrices, themselves, are not too compelling, unless there is a relationship between the row and column variables
Square-symmetric matrix Any matrix that has the same number of rows and columns (e.g., $n \times n$), plus has symmetry above and below the diagonal.	$\mathbf{S} = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 7 & 10 \\ 1 & 10 & 0 \end{bmatrix}$	The transpose of this matrix produces the exact same matrix
Diagonal matrix Only the elements along the **diagonal** of a matrix have value. All off-diagonal elements are 0.	$\mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 5 \end{bmatrix}$	It should be obvious that this matrix needs to be a square-symmetric matrix.
Identity matrix A diagonal matrix that has all diagonal elements equal to one	$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	This is the matrix equivalent to the value 1. For example, the product **AI** is equal to **A**. (Note that **A** does not need to be square or symmetric, but only have the same number of columns as **I** has rows.)

Orthogonal matrix A square-symmetric matrix that has the property that the product of it and its transpose produces an identify matrix. I.e., $\mathbf{AA}^t = \mathbf{I}$	$\mathbf{A} = \begin{bmatrix} .7071 & .7071 \\ .7071 & -.7071 \end{bmatrix}$	$\mathbf{AA}' = \begin{bmatrix} .7071 & .7071 \\ .7071 & -.7071 \end{bmatrix} \begin{bmatrix} .7071 & .7071 \\ .7071 & -.7071 \end{bmatrix}$ $\mathbf{AA}' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$
Ones matrix A matrix or vector (more typical) of only ones. (A **Zeros** matrix has all 0s.)	$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$	A ones vector is rather useful in statistics calculations. For example, for any ones vector of length n, $\mathbf{1}^t\mathbf{1} = n$. This might seem trivial, but it can be useful for calculating means or finding y-intercepts in linear regression (which will become apparent later.)
Matrix inverse Provided a matrix is square-symmetric, it is invertible if, and only if, the product $\mathbf{AA}^{-1} = \mathbf{I}$.	$\mathbf{AA}^{-1} = \mathbf{I}$	A demonstration of this principal is given below. This will be important for understanding linear models, using matrix calculations.

Multiplying data and other matrices has geometric interpretations

YI=**Y**: No change to **Y**

Y*c***I**=$\mathbf{Y}_{sc}$: Change of scale (e.g, enlargement)

YD=$\mathbf{Y}_{stretch}$: Stretching if **D** is diagonal

YT=$\mathbf{Y}_{rot}$: Rigid rotation if **T** is *p*x*p* orthogonal

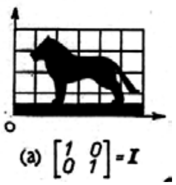
YT=$\mathbf{Y}_{new}$: Shear if **T** is not orthogonal

(T can be decomposed into rotation, dilation, rotation)

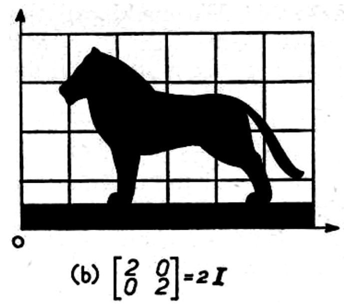
Y = data matrix

n rows and 2 columns

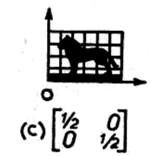
$$\mathbf{H} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



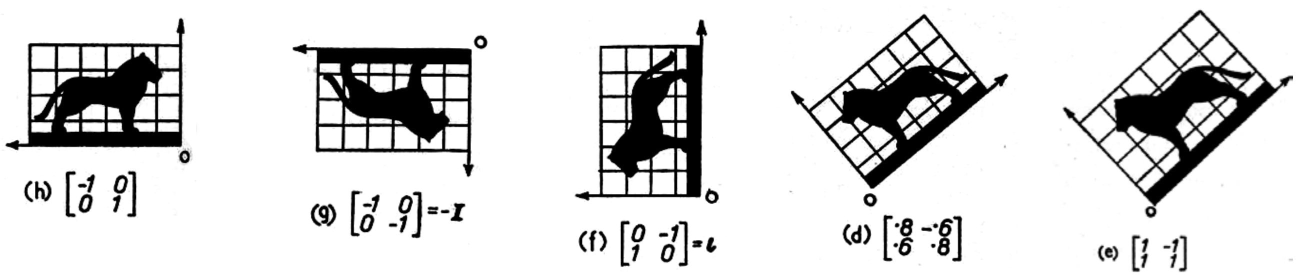
Original



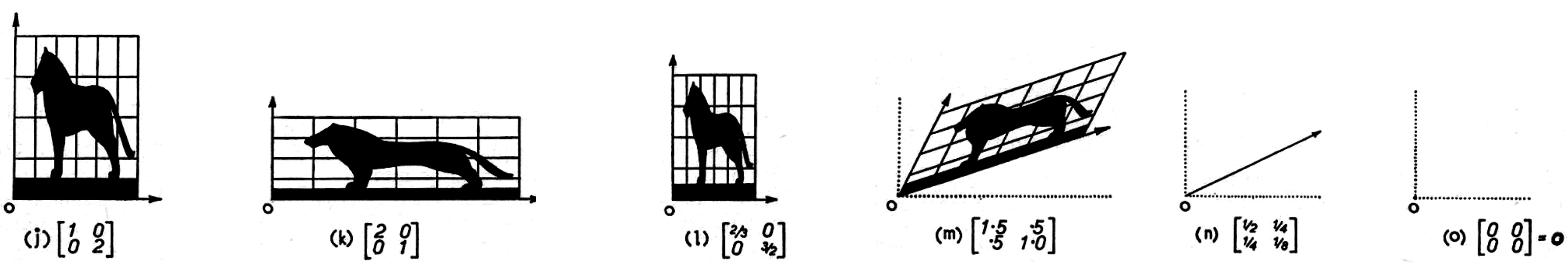
Scalar (2)



Scalar (1/2)



Rotations



Shears and Projections

(images from C.A.B. Smith, 1969)

Matrices cannot be divided.

-A matrix inverse (reciprocal) may be identified, then multiply by inverse

-Only some matrices can be inverted

-Inverses have the property: $\mathbf{AA}^{-1}=\mathbf{I}$

Matrix needs to be square and non-singular (determinant ≠ 0)

$$\mathbf{A}=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

The matrix inverse is found as

$$\mathbf{A}^{-1}=\frac{1}{\det(\mathbf{A})}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}=\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Matrix is singular if $ad-bc = 0$

Example

$$\mathbf{A} = \begin{bmatrix} 3 & 4 \\ 4 & 6 \end{bmatrix}$$

Inverse

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\begin{bmatrix} 6 & -4 \\ -4 & 3 \end{bmatrix} = \frac{1}{18-16}\begin{bmatrix} 6 & -4 \\ -4 & 3 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ -2 & 1.5 \end{bmatrix}$$

Confirm

$$\mathbf{AA}^{-1} = \begin{bmatrix} 3 & 4 \\ 4 & 6 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ -2 & 1.5 \end{bmatrix} = \begin{bmatrix} 3\cdot 3+4(-2) & 3(-2)+4\cdot 1.5 \\ 4\cdot 3+6(-2) & 4\cdot(-2)+6\cdot 1.5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Note, inversion with 3 × 3 matrices or larger, not a trivial exercise

Plagiarized from Wikipedia….

Inversion of 3x3 matrices [edit]

A computationally efficient 3x3 matrix inversion is given by

$$\mathbf{A}^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & I \end{bmatrix}^{T} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} A & D & G \\ B & E & H \\ C & F & I \end{bmatrix}$$

where the determinant of **A** can be computed by applying the rule of Sarrus as follows:

$$\det(\mathbf{A}) = a(ei - fh) - b(id - fg) + c(dh - eg).$$

If the determinant is non-zero, the matrix is invertible, with the elements of the above matrix on the right side given by

$$\begin{array}{ccc} A = (ei - fh) & D = -(bi - ch) & G = (bf - ce) \\ B = -(di - fg) & E = (ai - cg) & H = -(af - cd) \\ C = (dh - eg) & F = -(ah - bg) & I = (ae - bd) \end{array}$$

The Cayley–Hamilton decomposition gives

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \left[\frac{1}{2} \left((\mathrm{tr}\mathbf{A})^2 - \mathrm{tr}\mathbf{A}^2 \right) - \mathbf{A}\,\mathrm{tr}\mathbf{A} + \mathbf{A}^2 \right].$$

Computer programs like R use LAPACK – a system of numerical linear equations and decompositions – to estimate matrix inverse.

Matrix inversion used frequently as an analytical tool in statistical estimation

Consider:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{bmatrix}$$

$$\mathbf{X}^{t}\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 21 \\ 21 & 91 \end{bmatrix}$$

This is the matrix equivalent to squaring!

$$\left(\mathbf{X}^{t}\mathbf{X}\right)^{-1} = \frac{1}{6 \cdot 91 - 21 \cdot 21} \begin{bmatrix} 91 & -21 \\ -21 & 6 \end{bmatrix} = \frac{1}{105} \begin{bmatrix} 91 & -21 \\ -21 & 6 \end{bmatrix} = \begin{bmatrix} 0.876 & -0.200 \\ -0.200 & 0.867 \end{bmatrix}$$

X contains a dummy variable, and a second variable

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{bmatrix}$$

$$\begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix}$$

$$\mathbf{X}^{\mathrm{t}}\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 21 \\ 21 & 91 \end{bmatrix}$$

X is a linear model design matrix. It allows calculation of statistics for variables

- Here is the simplest linear model design matrix

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{x}^{\mathrm{T}}\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 5 = n \qquad \left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1} = \frac{1}{5} = \frac{1}{n}$$

- Here is a variable, *Y*, expressed as a vector

$$\mathbf{y} = \begin{bmatrix} 8 \\ 3 \\ 4 \\ 4 \\ 1 \end{bmatrix} \qquad \mathbf{x}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 8 \\ 3 \\ 4 \\ 4 \\ 1 \end{bmatrix} = 20 = \sum y$$

$$\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{y} = \frac{\sum y}{n} = \bar{y} = 4$$

The linear model and its coefficients

THIS IS THE MEAN

NULL MODEL

- Here is the linear model

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 8 \\ 7 \\ 4 \\ 4 \\ 1 \end{bmatrix} \quad \mathbf{x}^{\mathrm{T}}\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 5 = n \quad \left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1} = \frac{1}{5} = \frac{1}{n}$$

$$\mathbf{x}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 8 \\ 3 \\ 4 \\ 4 \\ 1 \end{bmatrix} = 20 = \sum y$$

$$\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{y} = \frac{\sum y}{n} = \bar{y} = 4$$

$$\mathbf{x}\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} \bar{y} \\ \bar{y} \\ \bar{y} \\ \bar{y} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \\ 4 \end{bmatrix}$$

$$\underset{n \times p}{\mathbf{Y}} = \underset{n \times k}{\mathbf{X}}\,\underset{k \times p}{\mathbf{B}} + \underset{n \times p}{\mathbf{E}}$$

Matrix of *p* dependent values

Model design matrix *n* subjects and *k* parameters

Model coefficients for *k* parameters, ***p* times**

Matrix of unexplained values

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}$$

$$k \times p$$

Need to solve this

$$\hat{\mathbf{B}} = \left(\mathbf{X}^{\mathbf{t}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathbf{t}}\mathbf{Y}$$

Why?

$$\mathbf{Y} = \mathbf{XB}$$

$$\mathbf{X}^{\mathbf{T}}\mathbf{Y} = \mathbf{X}^{\mathbf{T}}\mathbf{XB}$$

$$\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{XB}$$

$$\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y} = \mathbf{IB}$$

Hmm, doesn't look like 'standard regression coefficient equations I know…

$$\mathbf{Y} = \mathbf{XB} + \varepsilon$$

where: $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ $\mathbf{X} = \begin{bmatrix} 1 & X_i \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$ and: $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$

1. Expand matrixes: $\hat{\mathbf{B}} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{Y} = \left(\begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix}\begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix}\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$

2. Begin rewrite: $\left(\begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix}\begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 & \cdots & 1 \\ X_1 & \cdots & X_n \end{bmatrix}\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \left(\begin{bmatrix} \sum \mathbf{1} = \mathbf{n} & \sum \mathbf{X} \\ \sum \mathbf{X} & \sum \mathbf{X}^2 \end{bmatrix}\right)^{-1}\begin{bmatrix} \sum \mathbf{Y} \\ \sum \mathbf{XY} \end{bmatrix}$

2. From before:

$$\hat{\mathbf{B}} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{Y} = \left(\begin{bmatrix} \sum \mathbf{1} = \mathbf{n} & \sum \mathbf{X} \\ \sum \mathbf{X} & \sum \mathbf{X}^2 \end{bmatrix}\right)^{-1} \begin{bmatrix} \sum \mathbf{Y} \\ \sum \mathbf{XY} \end{bmatrix}$$

3. Calculate inverse:

$$\left(\begin{bmatrix} \sum \mathbf{1} = \mathbf{n} & \sum \mathbf{X} \\ \sum \mathbf{X} & \sum \mathbf{X}^2 \end{bmatrix}\right)^{-1} = \left(\begin{bmatrix} \dfrac{\sum X^2}{n\sum X^2 - \left(\sum X\right)^2} & \dfrac{-\sum X}{n\sum X^2 - \left(\sum X\right)^2} \\ \dfrac{-\sum X}{n\sum X^2 - \left(\sum X\right)^2} & \dfrac{n}{n\sum X^2 - \left(\sum X\right)^2} \end{bmatrix}\right)$$

4. Multiply

$$\left(\begin{bmatrix} \dfrac{\sum X^2}{n\sum X^2 - \left(\sum X\right)^2} & \dfrac{-\sum X}{n\sum X^2 - \left(\sum X\right)^2} \\ \dfrac{-\sum X}{n\sum X^2 - \left(\sum X\right)^2} & \dfrac{n}{n\sum X^2 - \left(\sum X\right)^2} \end{bmatrix}\right) \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix} = \begin{bmatrix} \dfrac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - \left(\sum X\right)^2} \\ \dfrac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$\begin{bmatrix} \dfrac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - \left(\sum X\right)^2} \\ \dfrac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \beta_1 = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2}$$

Rearrange to: $b_{Y \cdot X} = \dfrac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$

Denominator:

1: $\sum\left(X - \bar{X}\right)^2 = \sum\left(X - \bar{X}\right)\left(X - \bar{X}\right)$

2: $\sum\left(X - \bar{X}\right)\left(X - \bar{X}\right) = \sum\left(X^2 - 2X\bar{X} + \bar{X}^2\right)$ (expand multiplication)

3: $\sum\left(X^2 - 2X\bar{X} + \bar{X}^2\right) = \sum X^2 - 2\bar{X}\sum X + n\bar{X}^2$ (separate sum, extract constant)

4: $\sum X^2 - 2\dfrac{\sum X}{n}\sum X + n\left(\dfrac{\sum X}{n}\right)^2 = \sum X^2 - \dfrac{\left(\sum X\right)^2}{n}$ (since: $\bar{X} = \dfrac{\sum X}{n}$)

5: $\sum X^2 - \dfrac{\left(\sum X\right)^2}{n} = n\sum X^2 - \left(\sum X\right)^2$ (multiply by n)

<u>NOTE: lots of work, but can get there from matrices!</u>

Now we have the regression parameters, need to assess significance

F-ratio is: SSM/SSE (with df corrections)

Need to calculate full and reduced model SS

Full model (contains all terms) $SS_F = \sum\left(Y - \hat{Y}\right)^2 = \left(\mathbf{Y} - \mathbf{XB}\right)^t\left(\mathbf{Y} - \mathbf{XB}\right)$

Reduced model ($X^{\#}$ has 1 less term in it) $SS_R = \left(\mathbf{Y} - \mathbf{X}^{\#}\mathbf{B}^{\#}\right)^t\left(\mathbf{Y} - \mathbf{X}^{\#}\mathbf{B}^{\#}\right)$

Significance based on:

$$F_s = \frac{\left({SS_R - SS_F}\big/{q - q'}\right)}{{SS_F}\big/{q'}}$$

General formula for SS of a model effect (say, A):

$$SS_A = \mathbf{B}_F^T\mathbf{X}_F^T\mathbf{Y} - \mathbf{B}_R^T\mathbf{X}_R^T\mathbf{Y}$$

But what is all of this? <u>**A comparison of models!**</u>

See Rencher, 1995: p. 359

A Tale of Two models:

$$\mathbf{X_R} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad \mathbf{X_F} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

We follow the same procedure for both models

1) Estimate coefficients $\hat{\mathbf{B}}_\mathbf{R} = \left(\mathbf{X}_\mathbf{R}^\mathbf{T}\mathbf{X}_\mathbf{R}\right)^{-1}\left(\mathbf{X}_\mathbf{R}^\mathbf{T}\mathbf{Y}\right)$ $\hat{\mathbf{B}}_\mathbf{F} = \left(\mathbf{X}_\mathbf{F}^\mathbf{T}\mathbf{X}_\mathbf{F}\right)^{-1}\left(\mathbf{X}_\mathbf{F}^\mathbf{T}\mathbf{Y}\right)$

2) Estimate predicted values $\hat{\mathbf{Y}}_\mathbf{R} = \mathbf{X}_\mathbf{R}\hat{\mathbf{B}}_\mathbf{R}$ $\hat{\mathbf{Y}}_\mathbf{F} = \mathbf{X}_\mathbf{F}\hat{\mathbf{B}}_\mathbf{F}$

3) Estimate Error $\hat{\mathbf{E}}_\mathbf{R} = \mathbf{Y} - \hat{\mathbf{Y}}_\mathbf{R}$ $\hat{\mathbf{E}}_\mathbf{F} = \mathbf{Y} - \hat{\mathbf{Y}}_\mathbf{F}$

4) SS $\boxed{\mathbf{S}_\mathbf{R} = \hat{\mathbf{E}}_\mathbf{R}^\mathbf{T}\hat{\mathbf{E}}_\mathbf{R}}$ $\boxed{\mathbf{S}_\mathbf{F} = \hat{\mathbf{E}}_\mathbf{F}^\mathbf{T}\hat{\mathbf{E}}_\mathbf{F}}$

The Data:

$$\mathbf{X} = \begin{bmatrix} 1 \\ 4 \\ 5 \\ 7 \\ 9 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$$

(for matrix form):

$$\mathbf{X}_{new} = \begin{bmatrix} 1 & 1 \\ 1 & 4 \\ 1 & 5 \\ 1 & 7 \\ 1 & 9 \end{bmatrix}$$

```
> lm(yreg~xreg)
Coefficients:
(Intercept)          xreg
     0.3152        0.5163
```

R speak for matrixx multiplication

```
> solve(t(xnewreg)%*%xnewreg)%*%t(xnewreg)%*%yreg
          [,1]
[1,] 0.3152174
[2,] 0.5163043
```

```
> anova(lm(yreg~xreg))

Df Sum Sq Mean Sq F value  Pr(>F)
xreg        1 9.8098  9.8098  154.71 0.00112 **
Residuals   3 0.1902  0.0634
```

```
From full model:
t(yreg-yhatreg)%*%(yreg-yhatreg)
          [,1]
[1,] 0.1902174
```

Same idea, but must use special X-matrix coding

-Recode *k* groups in *k-1* **dummy variables** (columns) of **X**

$$\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

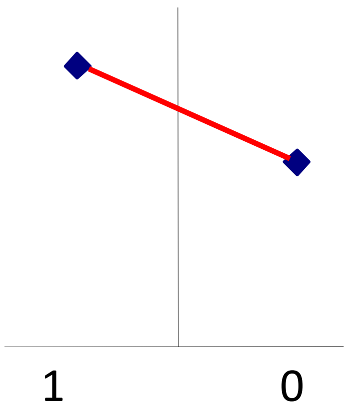
Generally, column 1 yields $\overline{\overline{\mathbf{X}}}$, column 2 yields deviation from $\overline{\overline{\mathbf{X}}}$ for mean of group 1, etc. (if set up as orthogonal contrasts)

Dummy variables can be constructed as 0/1 designators, OR as orthogonal contrasts (1/0/-1) of various groups

For GLM ANOVA, value in b represent the overall mean ($\bar{\bar{X}}$) and deviations from the grand mean ($\bar{\bar{X}} - \bar{X}_1$)

For 2 groups, think of b1 as the 'slope' between group means
Example:

$$\bar{\bar{X}} = 5$$
$$\bar{X}_1 = 6$$
$$\bar{X}_6 = 4$$



$$\beta_0 = 6$$
$$\beta_1 = -2$$

The Data:

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} 5 \\ 4 \\ 4 \\ 4 \\ 3 \\ 7 \\ 5 \\ 6 \\ 6 \\ 6 \end{bmatrix}$$

$$\overline{\overline{X}} = 5 \qquad \overline{X}_1 = 4 \quad \overline{X}_2 = 6$$

n_1=5

n_2=5

```
> summary(lm(y~x2))

Coefficients:
(Intercept)    6.0000      0.3162  18.974 1.69e-08
x2            -2.0000      0.4472  -4.472  0.00208

> solve(t(xnew)%*%xnew)%*%t(xnew)%*%y
     [,1]
[1,]    6
[2,]   -2
```

```
> anova(lm(y~x2))
Df Sum Sq Mean Sq F value   Pr(>F)
x2         1     10    10.0      20 0.002077 **
Residuals  8      4     0.5
```

From full model:

```
> t(y-yhat)%*%(y-yhat)
     [,1]
[1,]    4
```

GLM is the fitting of models

-Fit null model (e.g., H_0 = Y~1) and then more complicated models
-Assess fit of different models via SS (i.e,. LRT)

ANOVA and regression all the same model (GLM)
(key difference is what is used in X)

GLM parameters found using matrix algebra

$$\mathbf{B} = \left(\mathbf{X}^t \mathbf{X}\right)^{-1} \mathbf{X}^t \mathbf{Y}$$

Question: What about multivariate-Y?