20. Matrix Approach to Linear Regression

Readings: Kleinbaum, Kupper, Nizam, and Rosenberg (KKNR): Appendix B
https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/

SAS: PROC IML

Homework: Bonus Matrix Algebra Homework due by midnight on November 14
Homework 9 due by 11:59 pm on November 28 (Thanksgiving)

Overview

A) Re/Preview of Topics
B) Linear Algebra Review
C) Least Squares Estimation (LSE) Using Matrices
D) Matrix Example in SAS
E) Appendix with Corresponding R Code

A. Review (Lecture 19)/ Current (Lecture 20)/ Preview (Lecture 21)

Lecture 19:

- Correlation in terms of simple linear regression
- Multiple linear regression model (NOT multivariate regression)
- Comparing full and reduced models
  - Overall F-test (all β's=0, except the intercept)
  - Partial F-test (some β's=0)
  - t-statistic (one β=0)

Lecture 20:

- Linear Algebra Review
- Least Squares Estimation (LSE) using matrices

Lecture 21:

- Confounding X←C→Y
- Mediation X→M→Y

B. Matrices

A **matrix** (plural: matrices) is a rectangular array of numbers. Examples of matrices are as follows:

$$\mathbf{A}=\begin{bmatrix}3 & 4\\ 2 & 2\end{bmatrix} \quad \mathbf{B}=\begin{bmatrix}1 & 0 & 0 & -1\end{bmatrix} \quad \mathbf{C}=\begin{bmatrix}1 & 2\\ 2 & 3\\ 0 & 1\end{bmatrix} \quad \mathbf{d}=\begin{bmatrix}2\end{bmatrix} \quad \mathbf{E}=\begin{bmatrix}2\\ 3\\ 1\end{bmatrix}$$

The **dimensions** of a matrix tell us how many rows and columns, respectively, the matrix has.

- Matrix **A** is a 2 × 2 matrix (2 rows and 2 columns)
- Matrix **B** is a 1 × 4 matrix (1 row and 4 columns)
- Matrix **C** is a 3 × 2 matrix (3 rows and 2 columns)
- Matrix **d** is a 1 × 1 matrix (scalar) (1 row and 1 column)
- Matrix **E** is a 3 × 1 matrix (3 rows and 1 column)

In general, an $r\times c$ matrix has r rows and c columns and it is customary to express this as $\mathbf{A}_{r\times c}$.

- $\mathbf{A}_{2\times 2}$ $\mathbf{B}_{4\times 1}$ $\mathbf{C}_{3\times 2}$ $\mathbf{E}_{3\times 1}$

$$\mathbf{A}=\begin{bmatrix}3 & 4\\ 2 & 2\end{bmatrix} \quad \mathbf{B}=\begin{bmatrix}1 & 0 & 0 & -1\end{bmatrix} \quad \mathbf{C}=\begin{bmatrix}1 & 2\\ 2 & 3\\ 0 & 1\end{bmatrix} \quad \mathbf{d}=\begin{bmatrix}2\end{bmatrix} \quad \mathbf{E}=\begin{bmatrix}2\\ 3\\ 1\end{bmatrix}$$

Special matrices include:

- A matrix with one row and one column is called a **scalar**. (**d** is a scalar)
- A matrix with only one row is called a **row vector**. (**B** is a row vector)
- A matrix with only one column is called a **column vector**. (**E** is a column vector)
- A **square matrix** is a matrix with equal row and column dimensions (i.e., $r = c$). (**A** is a square matrix)

An **element** of a matrix **A** is a_{ij}, referring to the element in the i^{th} row and the j^{th} column of **A.**

- For example, in matrix **C**, $c_{31} = 0$.

The usual convention is to name matrices with boldface letters and elements of a matrix by lowercase subscripted letters.

- **A**, **B**, **C**, **E**
- a_{21}, b_{13}, c_{31}, e_{31}

The elements of a vector can be singly subscripted.

- b_{13} vs b_3

Addition and subtraction of matrices

Addition (or subtraction) of two or more matrices: dimensions of the 2 matrices must be equal.

The sum has the same dimensions as each of the component matrices.
i.e. below, the dimension of **A**, **B**, and **C** are 2x2

Addition (or subtraction) of two matrices is obtained by simply adding (or subtracting) the corresponding elements of the matrices. For example:

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 3+1 & 4+3 \\ 2+2 & 2+4 \end{bmatrix} = \begin{bmatrix} 4 & 7 \\ 4 & 6 \end{bmatrix} = \mathbf{C}$$

The usual rules for addition and subtraction of numbers apply to addition and subtraction of matrices, namely:

1. Commutative Law for Addition: (**A**+**B** = **B**+**A**)

2. Associative Law for Addition: (**A**+**B**)+**C** = **A**+(**B**+**C**)

Matrix multiplication

First, assume that two matrices **A** and **B** have dimensions $n \times p$ and $p \times r$, respectively.

In matrix multiplication, the number of columns of the first member of the product must equal the number of rows in the second member of the product. (i.e. ($n \times p$) & ($p \times r$)).

The resulting matrix product will be of dimension $n \times r$.

Each element, c_{ij} of **C** = **AB** is obtained by summing the products of the i^{th} row of **A** by the j^{th} column of **B**. For example,

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

$$\mathbf{AB} = \begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} 2 & 1 & -5 \\ 4 & -3 & 1 \end{bmatrix} = \begin{bmatrix} 3(2)+4(4) & 3(1)-4(3) & 3(-5)+4(1) \\ 2(2)+2(4) & 2(1)-2(3) & 2(-5)+2(1) \end{bmatrix} = \begin{bmatrix} 22 & -9 & -11 \\ 12 & -4 & -8 \end{bmatrix}$$

Another example is:

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & -1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{AB} = \begin{bmatrix} 3 & 1 & -1 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3(2) & 1(1) & -1(1) \end{bmatrix} = 6$$

Matrix multiplication continued

As stated above, for matrix multiplication to be defined, the number of columns of the first member of the product must equal the number of rows in the second member of the product.

The only exception to this rule, however, is when a scalar is multiplied by a matrix, in which case the scalar is multiplied by each element of the matrix.

$$3\begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 3\times 3 & 3\times 4 \\ 3\times 2 & 3\times 2 \end{bmatrix} = \begin{bmatrix} 9 & 12 \\ 6 & 6 \end{bmatrix}$$

In general, **AB ≠ BA** (the Commutative Law doesn't hold). Often the dimensions don't work (i.e. $A_{4x2}\ B_{2x3} = C_{4x3}$ vs $B_{2x3}\ A_{4x2}$ not compatible), and even if they do we usually have different solutions:

$$\mathbf{AB} = \begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 3(1)+4(2) & 3(3)+4(4) \\ 2(1)+2(2) & 2(3)+2(4) \end{bmatrix} = \begin{bmatrix} 11 & 25 \\ 6 & 14 \end{bmatrix}$$

$$\mathbf{BA} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1(3)+3(2) & 1(4)+3(2) \\ 2(3)+4(2) & 2(4)+4(2) \end{bmatrix} = \begin{bmatrix} 9 & 10 \\ 14 & 16 \end{bmatrix}$$

Matrix multiplication continued

The Associative Law holds: $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.

The following rules of matrix arithmetic are also valid (assuming the sizes of the matrices are such that the indicated operations can be performed):

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

$\mathbf{A}(\mathbf{B}-\mathbf{C}) = \mathbf{AB}-\mathbf{AC}$ $(\mathbf{B}-\mathbf{C})\mathbf{A} = \mathbf{BA}-\mathbf{CA}$

$a(\mathbf{B}+\mathbf{C}) = a\mathbf{B}+a\mathbf{C}$ $a(\mathbf{B}-\mathbf{C}) = a\mathbf{B}-a\mathbf{C}$

$(a+b)\mathbf{C} = a\mathbf{C}+b\mathbf{C}$ $(a-b)\mathbf{C} = a\mathbf{C}-b\mathbf{C}$

$a(b\mathbf{C}) = (ab)\mathbf{C}$ $a(\mathbf{BC}) = (a\mathbf{B})\mathbf{C} = \mathbf{B}(a\mathbf{C})$

Transpose

The **transpose** of a matrix is obtained by writing each column of **A** as a row of $\mathbf{A}^T$.

- Denoted **A'** or $\mathbf{A}^T$
- For example:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 7 & 8 \\ 9 & 5 & 6 \end{bmatrix} \quad \text{then} \quad \mathbf{A}^T = \begin{bmatrix} 1 & 3 & 9 \\ 2 & 7 & 5 \\ 4 & 8 & 6 \end{bmatrix}$$

The transpose of a product $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ is the product of the transposes, in opposite order.

If **A** is a square matrix with $\mathbf{A} = \mathbf{A}^T$, then $a_{ij} = a_{ji}$ for all i and j, and **A** is a **symmetric matrix**.

- An example of a symmetric matrix:

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 7 & 5 \\ 4 & 5 & 6 \end{bmatrix} = \mathbf{A}^T$$

Examples of symmetric matrices in regression include *covariance matrices* and *correlation matrices*.

Transpose cont.

The transpose of a column vector is a row vector.

The transpose of a row vector is a column vector.

Suppose that **A** is an $n \times 1$ vector with elements $a_1, a_2, \ldots, a_n$:

- Then the product $\mathbf{A}^T\mathbf{A}$ is well defined, and it is the 1×1 scalar. For example,

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a_1^2 + a_2^2 + \cdots + a_n^2 = \sum a_i^2$$

- $\mathbf{A}^T\mathbf{A}$ is the sum of squares of the elements of the vector **A**.

- The square root of this quantity is called the **norm** or **length** of the vector **A** and is given by $||\mathbf{A}||$, where $||\mathbf{A}||^2 = \mathbf{A}^T\mathbf{A}$.

Diagonal matrices and trace

A square matrix is **diagonal** if all elements off the main diagonal are zero (i.e. $a_{ij} = 0$ unless $i = j$)

- An example of a diagonal matrix is:

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

For square matrices, the **trace** is the sum of the diagonal elements:

- For example:

$$tr(\boldsymbol{G}) = tr\left(\begin{bmatrix} 3 & 2 & 4 \\ 2 & 7 & 5 \\ 4 & 5 & 6 \end{bmatrix}\right) = 3 + 7 + 6 = 16$$

Identity matrix

The **identity matrix**, **I**, is a square diagonal matrix with the main diagonals set to 1.

- For example, the 3 × 3 identity matrix is

$$\mathbf{I}_{3x3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

An important property of the identity matrix:

$$\mathbf{A}_{r\times c}\mathbf{I}_{c\times c} = \mathbf{I}_{r\times r}\mathbf{A}_{r\times c} = \mathbf{A}_{r\times c}$$

- For example, if $\mathbf{A}=\begin{bmatrix} 3 & 2 & 4 \\ 2 & 7 & 5 \\ 4 & 5 & 6 \end{bmatrix}$, we can see that

$$\mathbf{AI} = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 7 & 5 \\ 4 & 5 & 6 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 7 & 5 \\ 4 & 5 & 6 \end{bmatrix} = \mathbf{A}$$

The identity matrix serves the same algebraic function for matrix multiplication that the number 1 serves for ordinary scalar multiplication.

Inverse of a matrix

The analog to division in matrix operations is obtained by calculating the inverse of a matrix

- The **inverse** of a square matrix **A** is denoted $\mathbf{A}^{-1}$
- Property of the inverse:

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

Properties of Inverses

- Not all matrices have inverses.
    - For real numbers only zero has no inverse
    - The inverse of any nonzero real number k is $1/k$

- An **invertible matrix** has exactly one inverse.

- If **A** and **B** are invertible matrices of the same size, then:
    - **AB** is invertible
    - $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Methods for finding $\mathbf{A}^{-1}$

There are several methods which can be used to find the inverse of a matrix:

- *augmentation* method
- the method of *determinants* and *cofactors*
- the method of *pivot* points.

In the case of a 2 × 2 matrix, a simple formula can be used to find the inverse (if one exists).

Let the matrix **A** = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then

- **A** is invertible if *ad* - *bc* ≠ 0

- If **A** in invertible, the inverse is given by the formula:

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

Methods for finding $\mathbf{A}^{-1}$ continued

For example, find the inverse of $\mathbf{A} = \begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix}$.

$$\mathbf{A}^{-1} = \frac{1}{3(2)-2(4)} \times \begin{bmatrix} 2 & -4 \\ -2 & 3 \end{bmatrix} = -\frac{1}{2} \times \begin{bmatrix} 2 & -4 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 1 & -\frac{3}{2} \end{bmatrix}$$

We can check if we calculated the correct inverse for **A** if $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$:

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 1 & -\frac{3}{2} \end{bmatrix} = \begin{bmatrix} 3(-1) + 4(1) & 3(2) - 4\left(\frac{3}{2}\right) \\ 2(-1) + 2(1) & 2(2) - 2\left(\frac{3}{2}\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Augmentation method for calculation $\mathbf{A}^{-1}$

To find the inverse of an invertible matrix **A**, we can find a sequence of elementary row operations that reduces **A** to the identity matrix and perform this same sequence of operations on $\boldsymbol{I}_n$ to obtain $\mathbf{A}^{-1}$.

To apply this method, we shall *augment* the original matrix by the identity matrix, thereby producing a matrix of the form:

$$[\ \mathbf{A} \mid \boldsymbol{I}\ ]$$

Then we shall apply row operations to the matrix until the left side is reduced to $\boldsymbol{I}$; these operations will convert the right side to $\boldsymbol{A}^{-1}$, so that the final matrix will have the form:

$$[\ \boldsymbol{I} \mid \mathbf{A}^{-1}\ ]$$

Example: Find the inverse of

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

(steps on the following slide)

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 3 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

Add -2 times row 1 to row 2

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -5 & -2 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

Multiply row 2 by -1

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 5 & 2 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

Add -1 times row 2 to row 3

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 5 & 2 & -1 & 0 \\ 0 & 0 & -4 & -2 & 1 & 1 \end{array}\right]$$

Add -2 times row 2 to row 1

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -7 & -3 & 2 & 0 \\ 0 & 1 & 5 & 2 & -1 & 0 \\ 0 & 0 & -4 & -2 & 1 & 1 \end{array}\right]$$

Add 5/4 times row3 to row 2

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -7 & -3 & 2 & 0 \\ 0 & 1 & 0 & -1/2 & 1/4 & 5/4 \\ 0 & 0 & -4 & -2 & 1 & 1 \end{array}\right]$$

Multiply row 3 by -1/4

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -7 & -3 & 2 & 0 \\ 0 & 1 & 0 & -1/2 & 1/4 & 5/4 \\ 0 & 0 & 1 & 1/2 & -1/4 & -1/4 \end{array}\right]$$

Add 7 times row 3 to row 1

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/2 & 1/4 & -7/4 \\ 0 & 1 & 0 & -1/2 & 1/4 & 5/4 \\ 0 & 0 & 1 & 1/2 & -1/4 & -1/4 \end{array}\right]$$

Inverse obtained! $\mathbf{A}^{-1} = \begin{bmatrix} 1/2 & 1/4 & -7/4 \\ -1/2 & 1/4 & 5/4 \\ 1/2 & -1/4 & -1/4 \end{bmatrix}$

Rank of a matrix

The **rank** of a matrix is the number of linearly independent columns or rows:

- If $n > p$, an $n \times p$ matrix **X** has **full rank** if the rank is p
- otherwise, **X** is **rank deficient**

A set of vectors is **linearly independent** if no vector in the set is:

1. A scalar multiple of another vector in the set, or
2. A linear combination of other vectors in the set

Conversely, a set of vectors is **linearly dependent** if any vector in the set is:

1. A scalar multiple of another vector in the set, or
2. A linear combination of other vectors in the set

Example: Consider **A** = $[1 \quad 2 \quad 3]$, **B** = $[4 \quad 5 \quad 6]$, **C** = $[5 \quad 7 \quad 9]$, **D** = $[2 \quad 4 \quad 6]$

- The set {**A**, **B**} are linearly independent because neither vector is a scalar multiple of the other.
- The set {**A**, **B**, **C**} are linearly dependent because **C** = **A** + **B** (a linear combination).
- The set {**A**, **D**} are linearly dependent because they are scalar multiples: **D** = 2**A**.

Rank of a matrix continued

A square matrix is of **full rank** or **nonsingular** if and only if the matrix has an inverse $\mathbf{A}^{-1}$

- This only occurs if the **determinant** of **A**, det(**A**), is not zero

The determinant of a 2×2 matrix can be calculated as:

$$det(\mathbf{A}) = det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = (a \times d) - (b \times c)$$

The determinant of square matrices larger than 2×2 is more complex to compute and methods have been developed to simplify their evaluation (e.g., reducing the matrix to row-echelon form)

Determinant for a 3 × 3 matrix: $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}\begin{matrix} 1 & 2 \\ 2 & 3 \\ 0 & 1 \end{matrix}$

$$det\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix} = (1 \times 3 \times 1) + (2 \times 1 \times 0) + (3 \times 2 \times 1) - (2 \times 2 \times 1) - (1 \times 1 \times 1) - (3 \times 3 \times 0) = 4$$

More examples of linearly dependent columns and/or rows

$det\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix} = 4 - 4 = 0$, since the determinant is 0 we know it isn't full rank (also 2x[1 2]=[2 4])

det(**A**) = $det\begin{bmatrix}1 & 2 & 3\\ 2 & 4 & 6\\ 5 & 5 & 5\end{bmatrix} = 1(4)5 + 2(6)5 + 3(2)5 - 3(4)5 - 6(5)1 - 5(2)2 = 0$

det($\mathbf{A}^T$) = $det\begin{bmatrix}1 & 2 & 5\\ 2 & 4 & 5\\ 3 & 6 & 5\end{bmatrix} = 1(4)5 + 2(5)3 + 2(6)5 - 5(4)3 - 6(5)1 - 5(2)2 = 0$

Augmentation: $\left.\begin{matrix}1 & 2 & 3\\ 2 & 4 & 6\\ 5 & 5 & 5\end{matrix}\right|\begin{matrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{matrix} \rightarrow \left.\begin{matrix}1 & 2 & 3\\ 0 & 0 & 0\\ 5 & 5 & 5\end{matrix}\right|\begin{matrix}1 & 0 & 0\\ -2 & 1 & 0\\ 0 & 0 & 1\end{matrix}$ multiply -2 times row 1 and add it to row 2

Augmentation: $\left.\begin{matrix}1 & 2 & 5\\ 2 & 4 & 5\\ 3 & 6 & 5\end{matrix}\right|\begin{matrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{matrix} \rightarrow \left.\begin{matrix}1 & 2 & 5\\ 0 & 0 & -5\\ 0 & 0 & -10\end{matrix}\right|\begin{matrix}1 & 0 & 0\\ -2 & 1 & 0\\ 0 & 0 & 1\end{matrix}$

First, multiply -2 times row 1 and add it to row 2
Second, multiply -3 times row 1 and add it to row 3

Orthogonal matrices

An $n \times n$ matrix $\mathbf{Q}$ is orthogonal if $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}_n$.

- $\mathbf{Q}^{-1} = \mathbf{Q}^T$
- For example, the matrix

$\mathbf{Q} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}$ is orthogonal, $\mathbf{Q}^T = \mathbf{Q}^{-1} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix}$

since $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{-1} = \mathbf{Q}^{-1}\mathbf{Q} = \mathbf{I}$:

$$\boldsymbol{Q}\boldsymbol{Q}^{\boldsymbol{T}} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

$$= \begin{bmatrix} \left(\frac{1}{\sqrt{3}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{6}}\right)^2 & \left(\frac{1}{\sqrt{3}}\right)^2 + 0 - 2\left(\frac{1}{\sqrt{6}}\right)^2 & \left(\frac{1}{\sqrt{3}}\right)^2 - \left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{6}}\right)^2 \\ \left(\frac{1}{\sqrt{3}}\right)^2 + 0 - 2\left(\frac{1}{\sqrt{6}}\right)^2 & \left(\frac{1}{\sqrt{3}}\right)^2 + 0 + \left(\frac{2}{\sqrt{6}}\right)^2 & \left(\frac{1}{\sqrt{3}}\right)^2 + 0 - 2\left(\frac{1}{\sqrt{6}}\right)^2 \\ \left(\frac{1}{\sqrt{3}}\right)^2 - \left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{6}}\right)^2 & \left(\frac{1}{\sqrt{3}}\right)^2 + 0 - 2\left(\frac{1}{\sqrt{6}}\right)^2 & \left(\frac{1}{\sqrt{3}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{6}}\right)^2 \end{bmatrix} = \boldsymbol{I}_{3x3}$$

Orthogonal matrices and vectors continued

Two $n \times n$ matrices **A** and **B** are orthogonal if **AB**=**BA**=0.

If **U** and **V** are $n \times 1$ column vectors, the **inner product** or **dot product** is $\mathbf{U} \cdot \mathbf{V} = \mathbf{U}^T\mathbf{V}$

- If $\mathbf{U}^T\mathbf{V}=0$, then **U** and **V** are orthogonal vectors
- Denoted $\mathbf{U} \perp \mathbf{V}$

Let $\mathbf{U}^T = [2 \quad 1 \quad 1 \quad 2]$ and $\mathbf{V}^T = [1 \quad 0 \quad 0 \quad -1]$, then

$$\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{V} = [2 \quad 1 \quad 1 \quad 2]\begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} = 2(1) + 1(0) + 1(0) - 2(1) = 0$$

$$\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{U} = [1 \quad 0 \quad 0 \quad -1]\begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \end{bmatrix} = 1(2) + 0(1) + \ 0(1) - 1(2) = 0$$

Eigenvectors and Eigenvalues

For a square *n*×*n* matrices **A**, eigenvector v is the non-zero vector and eigenvalue λ is the non-zero scalar such that $Av = \lambda v$.

For example:

$$Av_1 = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 2(1) + 0(0) - 1(1) \\ 0(1) + 3(0) - 1(0) \\ 1(1) + 0(0) - 2(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \lambda_1 v_1$$

$$Av_2 = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ c \\ 1 \end{bmatrix} = \begin{bmatrix} 2(1) + 0(c) + 1(1) \\ 0(1) + 3(c) + 1(0) \\ 1(1) + 0(c) + 2(1) \end{bmatrix} = \begin{bmatrix} 3 \\ 3c \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ c \\ 1 \end{bmatrix} = \lambda_2 v_2$$

$$v_1^T v_2 = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ c \\ 1 \end{bmatrix} = 1 - 1 = 0$$

Performing Matrix Manipulations in SAS with PROC IML

Addition:

$$\begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

Code

```
proc iml;
     A1 = {3 4 , 2 2};
/* row1, row2*/
     B1 = {1 3 , 2 4};
     C1 = A1 + B1;
     print C1;
quit;
```

Output

C1	
4	7
4	6

Multiplication:

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

Code

```
proc iml;
     A2 = {3 4 , 2 2};
     B2 = {2 1 -5 , 4 -3 1};
     A2B2 = A2 * B2;
     print A2B2;
quit;
```

Output

A2B2		
22	-9	-11
12	-4	-8

Transpose:

$$\begin{bmatrix} 1 & 2 & 4 \\ 3 & 7 & 8 \\ 9 & 5 & 6 \end{bmatrix}^T$$

Code

```
proc iml;
     A3 = {1 2 4 , 3 7 8 , 9 5 6};
     A3T = A3`;
     print A3T;
quit;
```

Output

A3T		
1	3	9
2	7	5
4	8	6

Trace:

$$\text{tr}\left(\begin{bmatrix} 3 & 2 & 4 \\ 2 & 7 & 5 \\ 4 & 5 & 6 \end{bmatrix}\right)$$

Code

```
proc iml;
     A4 = {3 2 4 , 2 7 5 , 4 5 6};
     trA4 = TRACE(A4);
     print trA4;
quit;
```

Output

trA4
16

Determinant:

$$\det\left(\begin{bmatrix} 3 & 2 & 4 \\ 2 & 7 & 5 \\ 4 & 5 & 6 \end{bmatrix}\right)$$

Code

```
proc iml;
     A4 = {3 2 4 , 2 7 5 , 4 5 6};
     detA4 = det(A4);
     print detA4;
quit;
```

Output

detA4
-5

Determinant and Inverse:

$$\det\left(\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}\right)$$

$$\& \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1}$$

Code

```
proc iml;
     A5 = {1 2 3 , 2 3 1 , 0 1 1};
     detA5 = det(A5);
     print detA5;
     invA5 = inv(A5);
     print invA5;
quit;
```

Output

detA5
4

invA5		
0.5	0.25	-1.75
-0.5	0.25	1.25
0.5	-0.25	-0.25

C. Linear regression in matrix notation

Matrix notation can be used to simplify many of the results used in multiple regression:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Where we define the matrices and vectors as:

$$\mathbf{Y}_{n\times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X}_{n\times(p+1)} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta}_{(p+1)\times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\epsilon}_{n\times 1} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Note, $\mathbf{X}$ is often called the **design matrix**.

Substituting in the definitions, we can also write Y_i in non-matrix format:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i$$

Linear regression in matrix notation

The sums of squares due to error can be written as:

$$SSE = SS_{Error} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

The least squares estimates are obtained by solving the following for $\boldsymbol{\beta}$

$$\frac{\partial SS_{Error}}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0$$

To get the least squares estimator of $\boldsymbol{\beta}$**,** rearrange and multiply each side by $(\mathbf{X}'\mathbf{X})^{-1}$ which gives

$$\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$$

$$(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y}$$

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y}$$

Assume $(\mathbf{X}'\mathbf{X})^{-1}$ exists, otherwise a generalized inverse (has some, but not all, properties of the ordinary inverse we defined earlier) can be used for a singular matrix

$(\mathbf{X}'\mathbf{X})^{-1}$ exists if the regressors are linearly independent (i.e., if no column of the **X** matrix is a linear combination of the other columns)

Simple Linear regression in matrix notation

For simple linear regression, **X** and **Y** are given by

$$\mathbf{Y}_{n\times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X}_{n\times 2} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$(\mathbf{X}'\mathbf{X}) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}, \qquad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}, \mathbf{Y}'\mathbf{Y} = \left(\sum y_i^2\right)$$

$(\mathbf{X}'\mathbf{X})^{-1}$ can be shown to be

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{nS_{XX}} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} = \frac{1}{S_{XX}} \begin{pmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

Then

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \frac{1}{S_{XX}} \begin{pmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} = \begin{pmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \frac{S_{XY}}{S_{XX}} \end{pmatrix}$$

Properties of the Least-Squares Estimators

$\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$ if $E(\boldsymbol{\varepsilon}) = \mathbf{0}$:

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon})] \\ &= E[\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon})] \\ &= \boldsymbol{\beta} \end{aligned}$$

By the Gauss-Markov theorem, $\widehat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$
https://en.wikipedia.org/wiki/Gauss%E2%80%93Markov_theorem

If we further assume that the errors are normally distributed:

- $\widehat{\boldsymbol{\beta}}$ is also the maximum-likelihood estimator of $\boldsymbol{\beta}$
    - We will show this in a later lecture

- Therefore, $\widehat{\boldsymbol{\beta}}$ is the minimum variance unbiased estimator (MVUE) of $\boldsymbol{\beta}$

Linear regression in matrix notation

The vector of fitted *y*-values, $\hat{\mathbf{Y}}$, corresponding to the observed y-values **Y** is

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{HY}.$$

The $n \times n$ matrix **H** is referred to as the ***hat matrix***.

- It maps the vector of observed values into a vector of fitted values

The residuals, $\boldsymbol{e} = \mathbf{Y} - \hat{\mathbf{Y}}$, can be written as

$$\boldsymbol{e} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}.$$

The error sums of square (SSE) is given by

$$SS_{Error} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$$

The regression (model) sums of squares is given by $SS_{Model} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\bar{Y}^2$.

The total sums of squares is given by: $SS_{Total} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\bar{Y}^2 = \mathbf{Y}'\mathbf{Y} - n\bar{Y}^2$

$SS_{total} = SS_{error} + SS_{model}$ (simple linear regression)

$$\frac{\partial}{\partial\beta_0}SS_{Error} = \frac{\partial}{\partial\beta_0}\left(\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2\right) = \sum_{i=1}^{n}-2\left(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i\right) = -2\sum_{i=1}^{n}\hat{e}_i = 0 \rightarrow \sum_{i=1}^{n}\hat{e}_i = 0 \quad \text{(Note 1)}$$

$$\frac{\partial}{\partial\beta_1}SS_{Error} = \frac{\partial}{\partial\beta_1}\left(\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2\right) = \sum_{i=1}^{n}-2X_i\left(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i\right) = -2\sum_{i=1}^{n}X_i\hat{e}_i = 0 \rightarrow \sum_{i=1}^{n}X_i\hat{e}_i = 0 \quad \text{(Note 2)}$$

$$SS_{Total} = \sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \sum_{i=1}^{n}\left((Y_i-\hat{Y}_i)+(\hat{Y}_i-\bar{Y})\right)^2 = \sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2+\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2+2\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)(\hat{Y}_i-\bar{Y})$$

We need to show $2\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)(\hat{Y}_i-\bar{Y}) = 0$:

$$\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)(\hat{Y}_i-\bar{Y}) = \sum_{i=1}^{n}\hat{e}_i(\hat{Y}_i-\bar{Y}) = \sum_{i=1}^{n}\hat{Y}_i\hat{e}_i - \bar{Y}\sum_{i=1}^{n}\hat{e}_i$$

Substituting in the definition of $\hat{Y}_i$:

$$\sum_{i=1}^{n}\hat{Y}_i\hat{e}_i - \bar{Y}\sum_{i=1}^{n}\hat{e}_i = \hat{\beta}_0\sum_{i=1}^{n}\hat{e}_i + \hat{\beta}_1\sum_{i=1}^{n}X_i\hat{e}_i - \bar{Y}\sum_{i=1}^{n}\hat{e}_i = \hat{\beta}_0(0)+\hat{\beta}_1(0)-\bar{Y}(0) = 0$$

Therefore:

$$SS_{Total} = \sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2+\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2 = SS_{Model}+SS_{Error}$$

$SS_{total} = SS_{error} + SS_{model}$ (multiple linear regression)

$$SS_{Total} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left((Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})\right)^2 = SS_{Model} + SS_{Error} + 2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})$$

Again, we need to show that $2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = 0$:

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = \sum_{i=1}^{n}\hat{e}_i(\hat{Y}_i - \bar{Y}) = \hat{\boldsymbol{e}}^T(\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}_n) = \hat{\boldsymbol{e}}^T(\mathbf{X}\hat{\boldsymbol{\beta}} - \bar{Y}\mathbf{1}_n) = 0 \text{ by notes 1 and 2 below}$$

Note 1: $\hat{\boldsymbol{e}}^T\bar{Y}\mathbf{1}_n = \bar{Y}\sum_{i=1}^{n}\hat{e}_i = 0$, where $\mathbf{1}_n = (1 \quad 1 \quad \cdots \quad 1)^T$ [i.e., $\mathbf{1}_n$ is a vector of 1's]

Note 2:
$$\hat{\boldsymbol{e}}^T\mathbf{X} = (\mathbf{Y} - \hat{\mathbf{Y}})^T\mathbf{X} = \mathbf{Y}^T\mathbf{X} - \hat{\mathbf{Y}}^T\mathbf{X} = \mathbf{Y}^T\mathbf{X} - (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y})^T\mathbf{X} = \mathbf{Y}^T\mathbf{X} - \mathbf{Y}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X} = \mathbf{Y}^T\mathbf{X} - \mathbf{Y}^T\mathbf{X} = 0$$

Matrix Algebra Note 1: $(\mathbf{X}^T)^{-1} = (\mathbf{X}^{-1})^T$ [i.e., the inverse of the transpose is equal to the transpose of the inverse]

Therefore, applying matrix algebra note 1:
$$((\mathbf{X}^T\mathbf{X})^{-1})^T = ((\mathbf{X}^T\mathbf{X})^T)^{-1} = (\mathbf{X}^T(\mathbf{X}^T)^T)^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}$$

We can also note that $\hat{\mathbf{Y}}^T\hat{\mathbf{Y}} = \hat{\mathbf{Y}}^T\mathbf{Y} = \mathbf{Y}^T\hat{\mathbf{Y}}$:

$$\begin{aligned}
\hat{\mathbf{Y}}^T\hat{\mathbf{Y}} \quad &= (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y})^T(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}) \\
&= (\mathbf{Y}^T\mathbf{X}[(\mathbf{X}^T\mathbf{X})^{-1}]^T\mathbf{X}^T)(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}) \\
&= (\mathbf{Y}^T\mathbf{X}[(\mathbf{X}^T\mathbf{X})^T]^{-1}\mathbf{X}^T)(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}) \\
&= (\mathbf{Y}^T\mathbf{X}[\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T)(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}) \\
&= \mathbf{Y}^T\mathbf{X}[\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \\
&= \mathbf{Y}^T\mathbf{X}[\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{Y} \\
&= \mathbf{Y}^T\mathbf{H}\mathbf{Y}
\end{aligned}$$

Therefore:

$$\hat{\mathbf{Y}}^T\hat{\mathbf{Y}} = \mathbf{Y}^T\mathbf{H}\mathbf{Y} = \mathbf{Y}^T\hat{\mathbf{Y}}$$

$$\hat{\mathbf{Y}}^T\hat{\mathbf{Y}} = \mathbf{Y}^T\mathbf{H}\mathbf{Y} = (\mathbf{H}\mathbf{Y})^T\mathbf{Y} = \hat{\mathbf{Y}}^T\mathbf{Y}$$

Note: $\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \mathbf{I}$, because $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$.

Variances and Covariances

Recall that if X is any random variable and a is any constant, then Var(aX) = a^{2}Var(X).

The matrix analog can be stated as follows:

$$\mathrm{Var}(\mathbf{AX}) = \mathbf{A}\mathrm{Var}(\mathbf{X})\mathbf{A}^T \quad \text{or} \quad \mathrm{Var}(\mathbf{AX}) = \mathbf{A}\,\boldsymbol{\Sigma}_\mathbf{X}\,\mathbf{A}^T$$

where **X** is any random vector and **A** is any compatible matrix with fixed values.

On the next slide, we show that if Var(**e**)=$\boldsymbol{I}\sigma^2_{Y|X}$, then

$$Var(\widehat{\boldsymbol{\beta}}) = \hat{\sigma}^2_{Y|X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

The variances of the regression coefficients are given by the diagonal elements of the matrix and the covariances by the off-diagonal elements.

The variance of a fitted value (i.e., the expected mean μ for a given value X=X_0) is given by:

$$Var(\mu_{Y|x_0}) = x_0^T[Var(\widehat{\boldsymbol{\beta}})]x_0 = \hat{\sigma}^2_{Y|X}x_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}x_0.$$

The variance of a future predicted value of Y for a given X_0 for an individual is given by:

$$Var(\hat{Y}|x_0) = \hat{\sigma}^2_{Y|X}[1 + x_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}x_0].$$

Variance of $\widehat{\beta}$

$$
\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}) &= Var[(\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}] \\
&= Var[(\boldsymbol{X'X})^{-1}\boldsymbol{X}'(\boldsymbol{X\beta}+\boldsymbol{\epsilon})] \\
&= Var[(\boldsymbol{X'X})^{-1}\boldsymbol{X}'(\boldsymbol{X\beta}) + (\boldsymbol{X'X})^{-1}\boldsymbol{X}'(\boldsymbol{\epsilon})] \\
&= Var[(\boldsymbol{X'X})^{-1}(\boldsymbol{X'X})\boldsymbol{\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X}'(\boldsymbol{\epsilon})] \\
&= Var[\boldsymbol{I\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X}'(\boldsymbol{\epsilon})] \\
&= Var[(\boldsymbol{X'X})^{-1}\boldsymbol{X}'(\boldsymbol{\epsilon})] \\
&= [(\boldsymbol{X'X})^{-1}\boldsymbol{X}']Var(\boldsymbol{\epsilon})[(\boldsymbol{X'X})^{-1}\boldsymbol{X}']' \\
&= Var(\boldsymbol{\epsilon})[(\boldsymbol{X'X})^{-1}\boldsymbol{X}'][(\boldsymbol{X'X})^{-1}\boldsymbol{X}']' \\
&= Var(\boldsymbol{\epsilon})[(\boldsymbol{X'X})^{-1}\boldsymbol{X}'][(\boldsymbol{X}')'((\boldsymbol{X'X})^{-1})'] \\
&= Var(\boldsymbol{\epsilon})(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}(\boldsymbol{X'X})^{-1} \\
&= Var(\boldsymbol{\epsilon})\boldsymbol{I}(\boldsymbol{X'X})^{-1} \\
&= \hat{\sigma}^2_{Y|X}(\boldsymbol{X'X})^{-1}
\end{aligned}
$$

D. SAS Example: Blood Pressure and Birthweight (Rosner)

Systolic blood pressure (SBP), birthweight (oz), and age (days) were measured for 16 infants. Our MLR model of $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ is:

$$\begin{pmatrix} 89 \\ 90 \\ 83 \\ 77 \\ 92 \\ 98 \\ 82 \\ 85 \\ 96 \\ 95 \\ 80 \\ 79 \\ 86 \\ 97 \\ 92 \\ 88 \end{pmatrix} = \begin{pmatrix} 1 & 135 & 3 \\ 1 & 120 & 4 \\ 1 & 100 & 3 \\ 1 & 105 & 2 \\ 1 & 130 & 4 \\ 1 & 125 & 5 \\ 1 & 125 & 2 \\ 1 & 105 & 3 \\ 1 & 120 & 5 \\ 1 & 90 & 4 \\ 1 & 120 & 2 \\ 1 & 95 & 3 \\ 1 & 120 & 3 \\ 1 & 150 & 4 \\ 1 & 160 & 3 \\ 1 & 125 & 3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{15} \\ e_{16} \end{pmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 16 & 1925 & 53 \\ 1925 & 236875 & 6405 \\ 53 & 6405 & 189 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} 1409 \\ 170350 \\ 4750 \end{pmatrix}$$

$$\mathbf{Y}'\mathbf{Y} = (124751)$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 3.3415265 & -0.021734 & -0.200517 \\ -0.021734 & 0.0001918 & -0.000406 \\ -0.200517 & -0.000406 & 0.0752777 \end{pmatrix}$$

Blood pressure and birthweight example cont.

The least squares solution for $\widehat{\boldsymbol{\beta}}$ is:

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \begin{pmatrix} 3.3415265 & -0.021734 & -0.200517 \\ -0.021734 & 0.0001918 & -0.000406 \\ -0.200517 & -0.000406 & 0.0752777 \end{pmatrix} \begin{pmatrix} 1409 \\ 170350 \\ 4750 \end{pmatrix} = \begin{pmatrix} 53.450 \\ 0.1256 \\ 5.8877 \end{pmatrix}$$

The variance-covariance matrix (dispersion matrix) for $\widehat{\boldsymbol{\beta}}$ is:

$$\begin{aligned} (\boldsymbol{X'X})^{-1}\hat{\sigma}^2_{Y|X} &= \begin{pmatrix} 3.3415265 & -0.021734 & -0.200517 \\ -0.021734 & 0.0001918 & -0.000406 \\ -0.200517 & -0.000406 & 0.0752777 \end{pmatrix} 6.14630 \\ &= \begin{pmatrix} 20.53801 & -0.13358 & -1.23244 \\ -0.13358 & 0.00118 & -0.00250 \\ -1.23244 & -0.00250 & 0.46268 \end{pmatrix} \\ &= \boldsymbol{\Sigma} \end{aligned}$$

where $\hat{\sigma}^2_{Y|X} = \frac{SS_{Error}}{n-p-1} = \frac{\boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y}}{n-p-1} = 6.14630$

Note, sometimes the variance of a matrix, Var(**X**), is denoted as $\boldsymbol{\Sigma}_{\mathbf{X}}$ or simply **Σ**.

Blood pressure and birthweight example cont.

To calculate the t-statistic for a given $\hat{\beta}$, say $\hat{\beta}_1$, we can create a vector to "pick off" this coefficient and its SE from our matrices.

For $\hat{\beta}_1$, we can use the vector **c** = (0 1 0):

$$\hat{\beta}_1 = \boldsymbol{c}\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 53.450 \\ 0.1256 \\ 5.8877 \end{pmatrix} = 0.1256$$

$$\begin{aligned} \widehat{Var}(\hat{\beta}_1) &= \boldsymbol{c}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{c}'\hat{\sigma}^2_{Y|X} \\ &= \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3.3415265 & -0.021734 & -0.200517 \\ -0.021734 & 0.0001918 & -0.000406 \\ -0.200517 & -0.000406 & 0.0752777 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} (6.14630) \\ &= \begin{pmatrix} -0.021734 & 0.0001918 & -0.000406 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} 6.14630 \\ &= (0.0001918)(6.14630) \\ &= 0.001179 \end{aligned}$$

Thus, $t = \dfrac{\boldsymbol{c}\widehat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{c}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{c}'\hat{\sigma}^2_{X|Y}}} = \dfrac{\widehat{\beta}_1}{\sqrt{\widehat{Var}(\widehat{\beta}_1)}} = \dfrac{0.1256}{\sqrt{0.001179}} = 3.657 \sim t_{16-2-1} = t_{13}$

SAS Matrix Example

```
*****************************************
* Example of MLR in SAS using matrices    *
* Birthweight & SBP Example Rosner        *
*****************************************;
DATA birth;
INPUT sbp weight age;
LABEL    sbp    = "SBP (mmHg)"
         weight = "Weight (oz)"
         age    = "Age (days)";
DATALINES;
89 135 3
90 120 4
83 100 3
77 105 2
92 130 4
98 125 5
82 125 2
85 105 3
96 120 5
95  90 4
80 120 2
79  95 3
86 120 3
97 150 4
92 160 3
88 125 3
;
RUN;

PROC REG DATA=birth;
    MODEL sbp = weight age / clb covb xpx; /* xpx prints matrix crossproducts */
RUN;
```

Blood pressure and birthweight example in SAS cont.

xpx (model crossproducts) output:

Model Crossproducts X'X X'Y Y'Y					
Variable	**Label**	**Intercept**	**weight**	**age**	**sbp**
Intercept	Intercept	16	1925	53	1409
weight	Weight (oz)	1925	236875	6405	170350
age	Age (days)	53	6405	189	4750
sbp	SBP (mmHg)	1409	170350	4750	124751

(X'X)

(X'Y)

(X'Y)'

(Y'Y)

Blood pressure and birthweight example in SAS cont.

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	591.03564	295.51782	48.08	<.0001
Error	13	79.90186	6.14630		
Corrected Total	15	670.93750			

Root MSE	2.47917	**R-Square**	0.8809
Dependent Mean	88.06250	**Adj R-Sq**	0.8626
Coeff Var	2.81524		

Parameter Estimates								
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**95% Confidence Limits**	
Intercept	Intercept	1	53.45019	4.53189	11.79	<.0001	43.65964	63.24074
weight	Weight (oz)	1	0.12558	0.03434	3.66	0.0029	0.05140	0.19976
age	Age (days)	1	5.88772	0.68021	8.66	<.0001	4.41823	7.35721

Covariance of Estimates				
Variable	**Label**	**Intercept**	**weight**	**age**
Intercept	Intercept	20.538014219	-0.13358058	-1.232439878
weight	Weight (oz)	-0.13358058	0.0011789747	-0.00249504
age	Age (days)	-1.232439878	-0.00249504	0.4626790419

$(\mathbf{X}'\mathbf{X})^{-1}s^2_{Y|X} = \boldsymbol{\Sigma}$

```
/* Now with matrix notation and PROC IML */
PROC IML;
X={
1 135 3,
1 120 4,
1 100 3,
1 105 2,
1 130 4,
1 125 5,
1 125 2,
1 105 3,
1 120 5,
1  90 4,
1 120 2,
1  95 3,
1 120 3,
1 150 4,
1 160 3,
1 125 3};
Y =
{89,90,83,77,92,98,82,85,96,95,80,79,86,97,92,88};
xpx = x`*x;
xpy = x`*y;
ypy = y`*y;
xpxi = INV(xpx);
betahat = xpxi*xpy;
ssm=betahat`*x`*y-SUM(Y)**2/NROW(y);
sse=(y-x*betahat)`*(y-x*betahat);
sst=ssm+sse;
sighat2 = sse/(NROW(y)-NROW(betahat));
varb = xpxi*sighat2;
cb1 = {0 1 0};
b1 = cb1*betahat;
varb1=cb1*xpxi*cb1`*sighat2;
seb1=SQRT(varb1);
tb1 = cb1*betahat/seb1;
pb1 = 2*PROBT(-ABS(tb1),(NROW(y)-NROW(betahat)));
```

(continued on the next panel)

```
cb2 = {0 0 1};
b2 = cb2*betahat;
varb2=cb2*xpxi*cb2`*sighat2;
seb2=SQRT(varb2);
tb2 = cb2*betahat/seb2;
pb2 = 2*PROBT(-ABS(tb2),(NROW(y)-NROW(betahat)));

PRINT y x;
PRINT xpx xpy ypy;
PRINT xpxi;
PRINT betahat;
PRINT ssm;
PRINT sse;
PRINT sst;
PRINT sighat2;
PRINT varb;
PRINT b1 seb1 tb1 pb1;
PRINT b2 seb2 tb2 pb2;
/* NOTE: You can change the field width for printing
numeric values with the RESET command */
RESET fw=12;
PRINT xpxi;
RESET fw=4;
PRINT b2 seb2 tb2 pb2;
```

Recall:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$Var\left(\widehat{\boldsymbol{\beta}}\right) = (\mathbf{X}'\mathbf{X})^{-1}\hat{\sigma}^2_{Y|X}$$

$$\hat{\sigma}^2_{Y|X} = \frac{SS_{Error}}{n-(p+1)}$$

$$SS_{Error} = \left(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}}\right)$$

$$SS_{Model} = \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\bar{Y}^2$$

Y	X		
89	1	135	3
90	1	120	4
83	1	100	3
77	1	105	2
92	1	130	4
98	1	125	5
82	1	125	2
85	1	105	3
96	1	120	5
95	1	90	4
80	1	120	2
79	1	95	3
86	1	120	3
97	1	150	4
92	1	160	3
88	1	125	3

xpx			xpy	ypy
16	1925	53	1409	124751
1925	236875	6405	170350	
53	6405	189	4750	

xpxi		
3.3415265	-0.021734	-0.200517
-0.021734	0.0001918	-0.000406
-0.200517	-0.000406	0.0752777

betahat
53.450194
0.1255833
5.8877191

ssm
591.03564

ssm
591.03564

sst
670.9375

sighat2
6.1462969

varb		
20.538014	-0.133581	-1.23244
-0.133581	0.001179	-0.002495
-1.23244	-0.002495	0.462679

b1	seb1	tb1	pb1
0.1255833	0.0343362	3.6574594	0.0028958

b2	seb2	tb2	pb2
5.8877191	0.6802051	8.6557992	9.3419E-7

xpxi		
3.3415265201	-0.021733506	-0.200517464
-0.021733506	0.0001918187	-0.000405942
-0.200517464	-0.000405942	0.075277691

b2	seb2	tb2	pb2
5.89	0.68	8.66	0

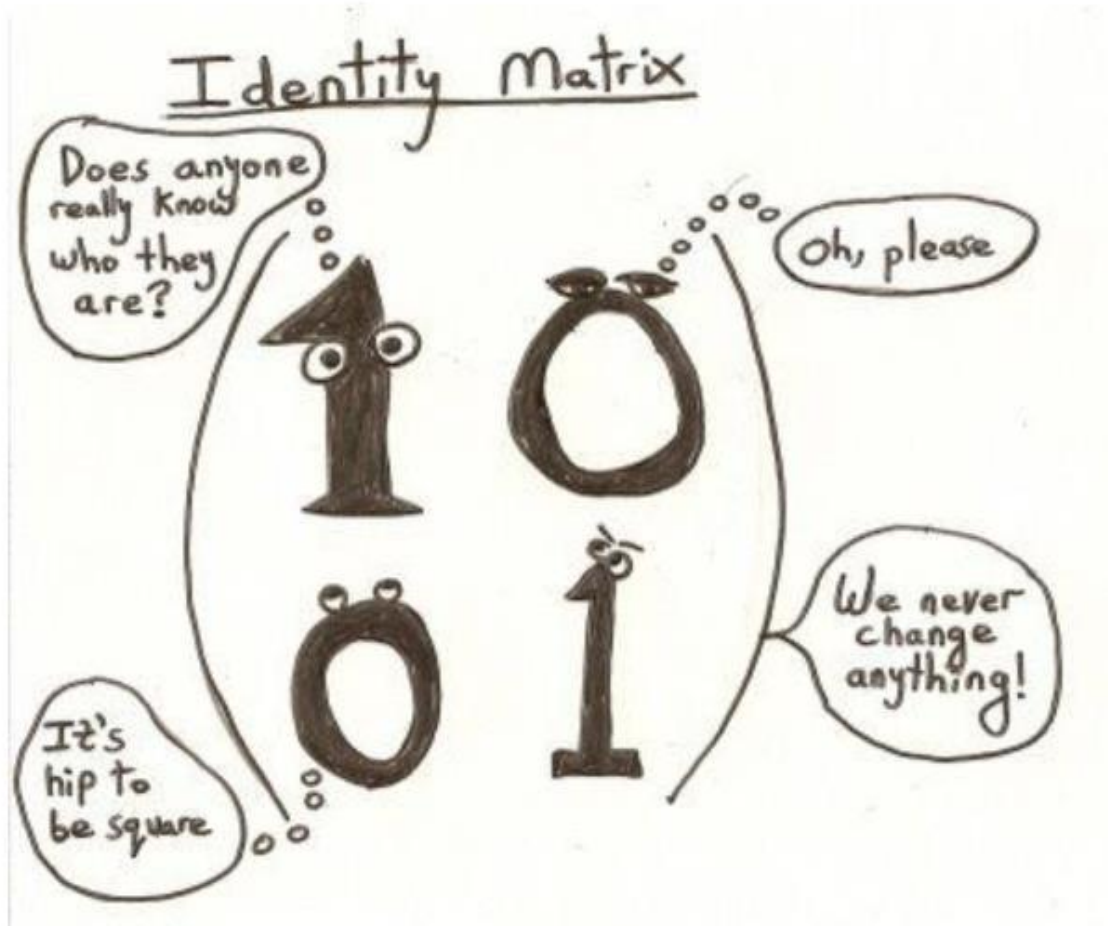
E. R Code for Matrix Calculations

```
### Lecture 20 Related Code: Matrix Approach to Linear Regression

# Unless otherwise noted, functions are included in base R and should not need
additional packages.

############################################################
### Anatomy of the matrix function

matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)

data: data vector to turn into a matrix

nrow, ncol:  indicates the number of rows and/or columns (if there is not enough
data, R will start "repeating" what is there to fill it all in)
  • if you only set one, it will try to determine the other based on the amount of
    data

byrow:  should the data be filled in column-by-column (FALSE) or row-by-row (TRUE)

dimnames:  you can name the rows and columns if it would be informative, provide a
list object:  list( rownames, colnames )
  • note, if only one set of names is provided, list( names ), it defaults to row
    names
```

```
############################################################
### General matrix manipulations (slides 24-25)

## Create matrices for use in examples on slides 24-25

mat1 <- matrix( c(3,4, 2,2), nrow=2, byrow=T) #2x2 matrix
mat2 <- matrix( c(1,3, 2,4), nrow=2, byrow=T) #2x2 matrix
mat3 <- matrix( c(2,1,-5, 4,-3,1), nrow=2, byrow=T) #2x3 matrix
mat4 <- matrix( c(1,2,4, 3,7,8, 9,5,6), nrow=3, byrow=T) #3x3 matrix
mat5 <- matrix( c(3,2,4, 2,7,5, 4,5,6), nrow=3, byrow=T) #3x3 matrix
mat6 <- matrix( c(1,2,3, 2,3,1, 0,1,1), nrow=3, byrow=T) #3x3 matrix

## Addition (uses the normal + operator)
mat1 + mat2

## Multiplication (uses %*% to indicate matrix multiplication)
mat1 %*% mat3

## Transpose (uses t() function)
t(mat4)

## Trace (can use the tr() function from the psych package or the matrix.trace() function from
the matrixcalc package)
library(psych)
library(matrixcalc)

tr(mat5) #from pysch package
matrix.trace(mat5) #from matrixcalc package

## Determinant (uses det() function)
det(mat5)

## Determinant and inverse (inverse uses solve() function)
det(mat6)
solve(mat6)
# In some cases, you may need to use the ginv() function from the MASS package to calculate the
Moore-Penrose generalized inverse
library(MASS)
ginv(mat6)
```

```
############################################################
### Blood Pressure and Birthweight from Rosner (slides 36-38, SAS code on slides 39-44)

Y <- matrix( c(89,90,83,77,92,98,82,85,96,95,80,79,86,97,92,88), ncol=1 )
X <- matrix( c( rep(1, length(Y)),
    c(135,120,100,105,130,125,125,105,120,90,120,95,120,150,160,125),
    c(3,4,3,2,4,5,2,3,5,4,2,3,3,4,3,3)),
    ncol=3, byrow=F)

###Slide 36 (setting up the matrix pieces)

##X'X (Xtranspose times X)
XtX <- t(X) %*% X
XtX

##X'Y (Xtranspose times Y)
XtY <- t(X) %*% Y
XtY

##Y'Y (Ytranspose times Y)
YtY <- t(Y) %*% Y
YtY

##(X'X)^(-1) ( (Xtranspose times X)inverse )
XtX_inv <- solve(XtX)
```

```
###Slide 37 (applying the matrix pieces to calculate our estimated beta values and their
covariance matrix)

## Calculate our beta coefficients
beta_vec <- XtX_inv %*% t(X) %*% Y
beta_vec

## Calculate the MSE (i.e., the variance of Y|X)
var_ygivenx <- ( YtY - t(beta_vec) %*% t(X) %*% Y ) / ( length(Y) - 2 - 1 ) #denominator is n-p-
1, where n=16, p=2 (2 different predictors)

## Calculate the covariance matrix for our betas
Sigma_mat <- XtX_inv * as.numeric( var_ygivenx ) #note: here we had to convert our sigma^2(Y|X)
into a numeric from a 1x1 matrix object, otherwise R gives a "Error...non-conformable arrays"


###Slide 38 (testing a single covariate for beta_1)

## Set up a vector to extract beta_1 from our betas
c_vec <- matrix(c(0,1,0), nrow=1)

## Extract beta_1
beta_1 <- c_vec %*% beta_vec
beta_1

## Calculate the variance for beta_1
var_beta_1 <- c_vec %*% XtX_inv %*% t(c_vec) * as.numeric(var_ygivenx)
var_beta_1

## Calculate the t-statistic
t_stat <- beta_1 / sqrt( var_beta_1 )
t_stat

## Calculate the p-value for our two-sided t-test (note, this isn't shown on slide 38, but is
helpful to know)
2 * (1-pt( t_stat, df = length(Y) - 2 - 1) ) # df = n-p-1

## Compare results to linear regression output
summary( lm( Y ~ X[,2:3] ) )
```


Identity Matrix
Does anyone really know who they are?
Oh, please
We never change anything!
It's hip to be square