Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

MATRIX ALGEBRA REVIEW FOR STATISTICS

LESSON 2 NOTES

Rev. 2.5

by

Robert A. LaBudde, Ph.D.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

VI. MATRIX INVERSE AND SOLUTION OF A SET OF LINEAR EQUATIONS:

Consider square matrices of order n x n. They can be added and subtracted, multiplied by scalars, multiplied times each other, and the operations are associative and distributive, although matrix multiplication is not commutative. There is an identity matrix for addition (i.e., the all zero matrix), so

$$\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A} \qquad \text{(VI.1)}$$

and there is an identity matrix for multiplication (**I**)

$$\mathbf{A}\,\mathbf{I} = \mathbf{I}\,\mathbf{A} = \mathbf{A} \qquad \text{(VI.2)}$$

All that is required for square matrices to fulfill the requirements for an “group algebra” (actually “group ring”, because multiplication is not, in general, commutative) is that multiplication have an inverse so that “division” is defined.

In the case of scalars, a multiplicative inverse and division are defined only if the scalar is nonzero (i.e., division is “nonsingular”). One might say all scalars are “invertible”, except the point 0, which has no inverse and causes division to be singular.

In the case of square matrices, which have two dimensions, more than just the zero matrix **0** is not invertible or singular. The 0 matrix can be thought to be zero in two dimensions, or a plane. A square matrix can also be “zero” in some sense as a line (one dimension). A matrix with one column or one row all zeroes is singular and noninvertible. So are other forms with similar hidden dependencies.

The key to finding the inverse of a square matrix **A** is to find the matrix **B** that reduces it to an identity matrix by multiplication, because then

$$\mathbf{A}\,\mathbf{B} = \mathbf{I} \qquad \text{(VI.3)}$$

If such a matrix **B** is found, it is unique, and usually denoted $\mathbf{A}^{-1}$. It is also true that it is a left as well as right inverse:

$$\mathbf{B}\,\mathbf{A} = \mathbf{I} \qquad \text{(VI.4)}$$

The matrix B can be found by solving the system of equations given in eq.(VI.3) or eq.(VI.4).

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

As an example, consider the matrix A = [1, 2; 3, 5]. It turns out its inverse is B = [-5, 2; 3, -1]. Using R,

```
> A = matrix(c(1,2, 3,5), ncol=2, byrow=TRUE) #2x2 matrix
> A
     [,1] [,2]
[1,]    1    2
[2,]    3    5

> B = solve(A)  #find inverse of A
> B
     [,1] [,2]
[1,]   -5    2
[2,]    3   -1

> A %*% B
              [,1] [,2]
[1,]  1.000000e+00    0
[2,] -8.881784e-16    1

> B %*% A
             [,1]          [,2]
[1,] 1.000000e+00 -1.554312e-15
[2,] 1.110223e-16  1.000000e+00
```

Note that eq.(VI.3) and eq.(VI.4) hold to the precision of the arithmetic used.

Some special cases have simple solutions.

For n = 1, the 1 x 1 matrix $\mathbf{A} = [\, A_{11} \,]$ has the (multiplicative) inverse $A^{-1} = [\, 1 / A_{11} \,]$, so long as $A_{11} \neq 0$. Otherwise $\mathbf{A}$ is singular and non-invertible.

For n = 2, the 2 x 2 matrix $A = [\, A_{11}, A_{12}; A_{21}, A_{22}]$ has the inverse

$$\mathbf{A}^{-1} \quad = \quad [\, A_{22}, -A_{12}; -A_{21}, A_{11} \,] / (A_{11} A_{22} - A_{12} A_{21}) \qquad \text{(VI.5)}$$

so long as the scalar quantity $A_{11} A_{22} - A_{12} A_{21} \neq 0$. Note that this condition is equivalent to an all-zero row or column, or one row is a multiple of the other, or one column is a multiple of the other.

For n > 2, the explicit formulas for $\mathbf{A}^{-1}$ become increasingly complicated, as does the condition for invertibility.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

The algorithm most commonly used for finding the inverse of a square matrix **A** is a process called "Gauss-Jordan reduction", based on the principle of "Gaussian elimination" used in solving systems of linear equations. This algorithm accomplishes its task of finding $\mathbf{A}^{-1}$ by carrying out methodically a sequence of elementary row operations, elementary column operations, or both.

Elementary row operations come in three "flavors". The first is the interchanging of two rows. The second is the multiplication of a row by a scalar. The third is subtracting one row from another. (The last two can be combined to subtract a multiple of one row from another.) Elementary column operations are defined similarly.

Elementary row operations are carried out by multiplying the target matrix **A** from the left by the row operator matrix **R**. Elementary column operations are carried out by right multiplying by the column operator matrix **C**.

For example, suppose **A** = [1, 2, 3; 4, 5, 6; 7, 8, 9]. To interchange rows 2 and 3, left multiply by **R** = [1, 0, 0; 0, 0, 1; 0, 1, 0]. To interchange columns 2 and 3, right multiply by **C** = [1, 0, 0; 0, 0, 1; 0, 1, 0]. (Note that these matrices are found by carrying out the same operations on the 3x3 identity matrix **I**.) Using R,

```
> A = matrix(c(1,2,3, 4,5,6, 7,8,9), ncol=3, byrow=TRUE) #3x3 matrix
> A
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8    9

> R = matrix(c(1,0,0, 0,0,1, 0,1,0), ncol=3, byrow=TRUE) #swap rows 2,3
> R
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    0    1
[3,]    0    1    0

> R %*% A #carry out operation
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    7    8    9
[3,]    4    5    6

> C = matrix(c(1,0,0, 0,0,1, 0,1,0), ncol=3, byrow=TRUE) #swap cols 2,3
> C
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    0    1
[3,]    0    1    0

> A %*% C #carry out operation
     [,1] [,2] [,3]
[1,]    1    3    2
[2,]    4    6    5
[3,]    7    9    8
```

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Suppose we wish to divide the second row by 2, or the second column by 2. Then, using R,

```
> A #show A again
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8    9

> R = matrix(c(1,0,0, 0,0.5,0, 0,0,1), ncol=3, byrow=TRUE) #divide row 2 by 2
> R
     [,1] [,2] [,3]
[1,]    1  0.0    0
[2,]    0  0.5    0
[3,]    0  0.0    1

> R %*% A #carry out operation
     [,1] [,2] [,3]
[1,]    1  2.0    3
[2,]    2  2.5    3
[3,]    7  8.0    9

> C = matrix(c(1,0,0, 0,0.5,0, 0,0,1), ncol=3, byrow=TRUE) #divide col 2 by 2
> C
     [,1] [,2] [,3]
[1,]    1  0.0    0
[2,]    0  0.5    0
[3,]    0  0.0    1

> A %*% C #carry out operation
     [,1] [,2] [,3]
[1,]    1  1.0    3
[2,]    4  2.5    6
[3,]    7  4.0    9
```

(Again note that **R** and **C** are found by applying the desired operation to **I**, and $\mathbf{C} = \mathbf{R}^t$.)

Finally, to subtract row 1 from row 3, or column 1 from column 3, then, using R:

```
> A #show A again
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8    9

> R = matrix(c(1,0,0, 0,1,0, -1,0,1), ncol=3, byrow=TRUE) #row 3 - row 1
> R
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]   -1    0    1

> R %*% A #carry out operation
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    6    6    6

> C = matrix(c(1,0,-1, 0,1,0, 0,0,1), ncol=3, byrow=TRUE) #col 3 - col 1
> C
```

```
     [,1] [,2] [,3]
[1,]    1    0   -1
[2,]    0    1    0
[3,]    0    0    1

> A %*% C #carry out operation
     [,1] [,2] [,3]
[1,]    1    2    2
[2,]    4    5    2
[3,]    7    8    2
```

The Gauss-Jordan algorithm for finding $\mathbf{A}^{-1}$ is to perform a sequence of row operations (or column operations, or both) on **A** until **A** matches the identity matrix **I**. Then the same sequence of row operations applied to **I** in the same order will result in $\mathbf{A}^{-1}$.

For example using R, for our matrix **A** = [1, 2, 3; 4, 5, 6; 7, 8, 9], first subtract 4 times row 1 from row 2 to make the element $A_{2\,1} = 0$ (actually a combination of elementary operations performed simultaneously):

```
> A #show A again
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8    9

> R1 = matrix(c(1,0,0, -4,1,0, 0,0,1), ncol=3, byrow=TRUE) #(row 2) - 4*(row 1)
> R1
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]   -4    1    0
[3,]    0    0    1

> B = R1 %*% A #carry out operation
> B  #result
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    0   -3   -6
[3,]    7    8    9
```

The resultant matrix **B** now has $a_{1\,1} = 1$ and $a_{2\,1} = 0$ in common with **I**. Now divide row 2 by –3 to make $A_{2\,2} = 1$:

```
> R2 = matrix(c(1,0,0, 0,-1/3,0, 0,0,1), ncol=3, byrow=TRUE) #(row 2)/(-3)
> R2
     [,1]       [,2] [,3]
[1,]    1  0.0000000    0
[2,]    0 -0.3333333    0
[3,]    0  0.0000000    1

> C = R2 %*% B #carry out operation
> C  #result
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    0    1    2
[3,]    7    8    9
```

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Now subtract 7 times row 1 from row 3 to make $A_{3\,1} = 0$:

```
> R3 = matrix(c(1,0,0, 0,1,0, -7,0,1), ncol=3, byrow=TRUE) #(row 3) - 7*(row 1)
> R3
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]   -7    0    1

> D = R3 %*% C #carry out operation
> D  #result
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    0    1    2
[3,]    0   -6  -12
```

Now add 6 times row 2 to row 3 to make $A_{3\,2} = 0$:

```
> R4 = matrix(c(1,0,0, 0,1,0, 0,6,1), ncol=3, byrow=TRUE) #(row 3) + 6*(row 2)
> R4
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    6    1

> E = R4 %*% D #carry out operation
> E  #result
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    0    1    2
[3,]    0    0    0
```

We have succeeded in row-reducing the original matrix **A** to an “upper-triangular” form, where all elements are zero below the “principal diagonal”, and this diagonal has all elements equal to one (“row echelon form”), except the last, which is 0. The next step would be to eliminate the three non-zero elements above the diagonal in a similar way, but we have run into a problem: The last row is all zero, instead of having a non-zero value for $A_{3\,3}$. The makes it impossible to get zeroes above it in the column. The all-zero row indicates that the matrix **A** is singular and non-invertible.

The reason **A** is singular is obvious, because twice the second row is equal to the sum of the first and third rows:

```
> 2*A[2,]
[1]  8 10 12

> A[1,] + A[3,]
[1]  8 10 12

> - A[1,] + 2*A[2,] - A[3,]   #linearly dependent rows
[1] 0 0 0
```

If there exist not all zero scalars c_1, c_2, c_3 such that

$$c_1 \mathbf{A}_{1\,.} + c_2 \mathbf{A}_{2\,.} + c_3 \mathbf{A}_{3\,.} = 0 \qquad \text{(VI.6)}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

where $\mathbf{A}_{i\,.}$ denotes the corresponding row vector of **A** for i = 1, 2, 3, then the rows of **A** are said to be "linearly dependent". Here, $c_1 = -1$, $c_2 = 2$, and $c_3 = -1$, so the rows of **A** are linearly dependent and therefore **A** is singular and non-invertible.

If we had been able to row reduce **A** to **I**, then the product of the row operation matrices in the order applied would give $\mathbf{A}^{-1}$. For our example, there is no inverse, but the combined operators do give the same combined result:

```
> R1to4 = R4 %*% R3 %*% R2 %*% R1   #combined operations
> R1to4
         [,1]       [,2] [,3]
[1,] 1.000000  0.0000000    0
[2,] 1.333333 -0.3333333    0
[3,] 1.000000 -2.0000000    1

> R1to4 %*% A       #combined result
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    0    1    2
[3,]    0    0    0
```

Row interchanges (called "pivoting" in algorithms) are usually also needed as the reduction occurs, although none were used in our example above. Typically, to avoid severe accumulation of round-off errors for larger matrices, rows are interchanged along the way to bring the largest element in the unreduced rows to the diagonal position. For example, if **A** = [0, 0, 1; 0, 1, 0; 1, 0, 0], rows 1 and 3 must be interchanged to make $\mathbf{A}_{1\,1} = 1$, as in the identity matrix:

```
> A = matrix(c(0,0,1, 0,1,0, 1,0,0), ncol=3, byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    0    0    1
[2,]    0    1    0
[3,]    1    0    0

> R1 = matrix(c(0,0,1, 0,1,0, 1,0,0), ncol=3, byrow=TRUE) #swap rows 1 and 3
> R1
     [,1] [,2] [,3]
[1,]    0    0    1
[2,]    0    1    0
[3,]    1    0    0

> R1 %*% A
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Note that, in this case, the single row operation matrix is actually $\mathbf{A}^{-1}$, as left multiplication by it results in the identity matrix **I**. Note that it is also a right inverse:

```
> A %*% R1
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

The fact that matrices behave like an algebra means that we can use their symbols in formulas in much that same way that we do scalars, and do formal solutions of equations by substituting matrix operations.

For example, consider a system of linear equations in unknowns x_1, x_2, ..., x_n:

$$
\begin{aligned}
A_{1\,1}\, x_1 + A_{1\,2}\, x_2 + \ldots + A_{1\,n}\, x_n &= b_1 \\
A_{2\,1}\, x_2 + A_{2\,2}\, x_2 + \ldots + A_{2\,n}\, x_n &= b_2 \\
\ldots & \\
A_{n\,1}\, x_1 + A_{n\,2}\, x_2 + \ldots + A_{n\,n}\, x_n &= b_n
\end{aligned}
\qquad \text{(VI.7)}
$$

These can be represented much more compactly and suggestively by

$$\mathbf{A}\,\mathbf{x} = \mathbf{b} \qquad \text{(VI.8)}$$

where **A** is the n x n matrix of coefficients $A_{i\,j}$, $\mathbf{x} = [x_1, x_2, ..., x_n]^t$, and $\mathbf{b} = [b_1, b_2, ..., b_n]^t$.

The form of eq.(VI.8) suggests how to obtain the solution by multiply both side by $\mathbf{A}^{-1}$ (thus “dividing” by **A**):

$$\mathbf{A}^{-1}\,\mathbf{A}\,\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b} \qquad \text{(VI.9a)}$$

$$\mathbf{I}\,\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b} \qquad \text{(VI.9b)}$$

$$\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b} \qquad \text{(VI.9c)}$$

This is clearly the solution of eq.(VI.8), because

$$\mathbf{A}\,(\mathbf{A}^{-1}\,\mathbf{b}) = (\mathbf{A}\,\mathbf{A}^{-1})\,\mathbf{b} = \mathbf{I}\,\mathbf{b} = \mathbf{b} \qquad \text{(VI.10)}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

EXAMPLE: CALCULATION OF INVERSE OF A 2 x 2 MATRIX

Q. If $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$*, what is* $\boldsymbol{A}^{-1}$*?*

A: If $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$, then

$$\begin{aligned} \mathbf{A}^{-1} &= \begin{bmatrix} 5 & -2 \\ -3 & 1 \end{bmatrix} / [(1)(5) - (2)(3)] \\ &= - \begin{bmatrix} 5 & -2 \\ -3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} \end{aligned}$$

Proof that $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$:

$$\begin{aligned} \mathbf{A}\,\mathbf{A}^{-1} &= \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad = \mathbf{I} \qquad\qquad \text{Q.E.D.} \end{aligned}$$

Using R:

```
> A<- matrix(c(1,2, 3,5), ncol=2, byrow=TRUE)
> A
     [,1] [,2]
[1,]    1    2
[2,]    3    5
> B<- solve(A) #find inverse
> B
     [,1] [,2]
[1,]   -5    2
[2,]    3   -1
> A %*% B   #prove inverse
              [,1] [,2]
[1,]  1.000000e+00    0
[2,] -8.881784e-16    1
> B %*% A   #prove inverse
             [,1]          [,2]
[1,] 1.000000e+00 -1.554312e-15
[2,] 1.110223e-16  1.000000e+00
```

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

APPENDIX: ALGORITHMIC ROW REDUCTION OF A MATRIX TO ITS INVERSE

Algorithm for row reduction of a matrix:

For k = 1, 2, ... , n: (i.e., all rows)

1. Find the largest modulus (absolute value) element in column k. Suppose it occurs in row m.

2. Interchange row k and row m to get the largest element on the diagonal.

3. Divide row k by the diagonal element to get 1.0 on the diagonal.

4. Subtract row k multiples from all other rows to make column k zero except on the diagonal.

Example:

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

Start by "augmenting" **A** with an identity matrix (set of columns) so that we can develop the inverse of **A** as we go along:

$$\begin{bmatrix} 0 & 2 & 4 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 2 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \text{->} \begin{bmatrix} 2 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 4 & 1 & 0 & 0 \end{bmatrix} \qquad \text{(Swap rows 1 and 3)}$$

$$\text{->} \begin{bmatrix} 1 & 0 & 1/2 & 0 & 0 & 1/2 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 4 & 1 & 0 & 0 \end{bmatrix} \qquad \text{(divide row 1 through by its diagonal element = 2)}$$

$$\text{->} \begin{bmatrix} 1 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & -1/2 & 0 & 1 & -1/2 \\ 0 & 2 & 4 & 1 & 0 & 0 \end{bmatrix} \qquad \text{(subtract row 1 from row 2 to zero the } A_{21} \text{ element)}$$

$$\text{->} \begin{bmatrix} 1 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 2 & 4 & 1 & 0 & 0 \\ 0 & 1 & -1/2 & 0 & 1 & -1/2 \end{bmatrix} \qquad \text{(swap rows 2 and 3 to get largest } A_{22} \text{ element)}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

-> $\begin{bmatrix} 1 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & 2 & 1/2 & 0 & 0 \\ 0 & 1 & -1/2 & 0 & 1 & -1/2 \end{bmatrix}$ (divide row 2 through by its diagonal element = 2)

-> $\begin{bmatrix} 1 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & 2 & 1/2 & 0 & 0 \\ 0 & 0 & -5/2 & -1/2 & 1 & -1/2 \end{bmatrix}$ (subtract row 2 from row 3 to zero A_{32})

-> $\begin{bmatrix} 1 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & 2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 1/5 & -2/5 & 1/5 \end{bmatrix}$ (divide row 3 through by its diagonal element = -5/2)

-> $\begin{bmatrix} 1 & 0 & 0 & -1/10 & 1/5 & 2/5 \\ 0 & 1 & 0 & 1/10 & 4/5 & -2/5 \\ 0 & 0 & 1 & 1/5 & -2/5 & 1/5 \end{bmatrix}$ (subtract multiples of row 3 from rows 1 and 2)

So rank(**A**) = 3 and the inverse of **A** is

$$\mathbf{A}^{-1} = \begin{bmatrix} -1/10 & 1/5 & 2/5 \\ 1/10 & 4/5 & -2/5 \\ 1/5 & -2/5 & 1/5 \end{bmatrix}$$

Using R:

```
> A<- matrix(c(0,2,4, 1,1,0, 2,0,1), ncol=3, byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    0    2    4
[2,]    1    1    0
[3,]    2    0    1

> solve(A)    #find inverse
     [,1] [,2] [,3]
[1,] -0.1  0.2  0.4
[2,]  0.1  0.8 -0.4
[3,]  0.2 -0.4  0.2
```

(Note the same process may be used to reduce **A** to row echelon form, except augmentation by an identity matrix is unnecessary as the inverse is not desired.)

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

VII. APPLICATION OF THE MATRIX INVERSE TO STATISTICS:

Consider the following small multivariate dataset on violent crime in ten major US cities, in rates of reported offenses per 100,000 population for a recent year:

City	Murder $\mathbf{x}_1$	Rape $\mathbf{x}_2$	Robbery $\mathbf{x}_3$	AggrAssault $\mathbf{x}_4$
Dallas	15.0	53.3	553.9	584.2
Houston	18.2	41.2	548.3	561.7
Las Vegas	11.6	54.6	409.0	507.7
Los Angeles	12.4	27.3	370.0	377.2
New York	7.3	13.1	287.9	329.6
Philadelphia	27.7	65.5	749.1	720.1
Phoenix	15.4	36.2	287.5	398.5
San Antonio	9.2	39.8	179.6	388.7
San Diego	5.4	27.5	170.8	300.8
San Jose	3.2	23.6	111.9	248.2

Consider the 10 x 4 matrix **A** formed by the numbers in this table.

First, find the mean values m = [m_1, m_2, m_3, m_4] of each column:

Average	12.54	38.21	366.8	441.67

Suppose we are interested in the correlation across cities of the incidence of column variable Murder ($\mathbf{x}_1$) and the incidence of the column variable Rape ($\mathbf{x}_2$) for the period involved:

$$r_{12} = \frac{\Sigma (x_{k1} - m_1)(x_{k2} - m_2)}{\sqrt{\{ [\Sigma (x_{k1} - m_1)^2] [\Sigma (x_{k2} - m_2)^2] \}}} \qquad k=1,2,...,10 \qquad \text{(VII.1)}$$

where the sum is over the cities and m_1 is the mean of first column and m_2 is the mean of the second column. The numerator in eq.(VII.1) should be recognized as the *inner* product of the two vectors ($\mathbf{x}_1 - m_1$) and ($\mathbf{x}_2 - m_2$), or

$$(\mathbf{x}_1 - m_1)^t (\mathbf{x}_2 - m_2) = \Sigma (x_{k1} - m_1)(x_{k2} - m_2) \qquad \text{(VII.2)}$$

Similarly, each of the sums in the denominator are the *inner* product of ($\mathbf{x}_1 - m_1$) and ($\mathbf{x}_2 - m_2$), resp., with themselves:

$$(\mathbf{x}_1 - m_1)^t (\mathbf{x}_1 - m_1) = \Sigma (x_{k1} - m_1)(x_{k1} - m_1) \qquad \text{(VII.3a)}$$

$$(\mathbf{x}_2 - m_2)^t (\mathbf{x}_2 - m_2) = \Sigma (x_{k2} - m_2)(x_{k2} - m_2) \qquad \text{(VII.3b)}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Therefore eq.(VII.1) can be written more succinctly in vector/matrix notation as

$$r_{12} = \frac{(\mathbf{x}_1 - m_1)^t (\mathbf{x}_2 - m_2)}{\sqrt{\{ [(\mathbf{x}_1 - m_1)^t (\mathbf{x}_1 - m_1)] [(\mathbf{x}_2 - m_2)^t (\mathbf{x}_2 - m_2)] \}}} \qquad \text{(VII.4a)}$$

$$= \frac{(\mathbf{x}_1 - m_1)^t (\mathbf{x}_2 - m_2)}{|\mathbf{x}_1 - m_1| \, |\mathbf{x}_2 - m_2|} \qquad \text{(VII.4b)}$$

Using R to perform the calculations involved:

```
> crime = read.csv('violent-crimes.csv', header=TRUE)
> crime
             Murder Rape Robbery AggrAssault
Dallas         15.0 53.3   553.9       584.2
Houston        18.2 41.2   548.3       561.7
Las Vegas      11.6 54.6   409.0       507.7
Los Angeles    12.4 27.3   370.0       377.2
New York        7.3 13.1   287.9       329.6
Philadelphia   27.7 65.5   749.1       720.1
Phoenix        15.4 36.2   287.5       398.5
San Antonio     9.2 39.8   179.6       388.7
San Diego       5.4 27.5   170.8       300.8
San Jose        3.2 23.6   111.9       248.2

> m = colMeans(crime)
> m
     Murder        Rape     Robbery AggrAssault
      12.54       38.21      366.80      441.67

> x1 = crime[,1]  #murder incidence
> x1
 [1] 15.0 18.2 11.6 12.4  7.3 27.7 15.4  9.2  5.4  3.2

> x2 = crime[,2]  #rape incidence
> x2
 [1] 53.3 41.2 54.6 27.3 13.1 65.5 36.2 39.8 27.5 23.6

> (x1 - m[1]) %*% (x2 - m[2]) #inner product
        [,1]
[1,] 787.326

> sqrt((x1-m[1]) %*% (x1-m[1])) #norm x1
         [,1]
[1,] 21.30315

> sqrt((x2-m[2]) %*% (x2-m[2])) #norm x2
         [,1]
[1,] 48.31448

> r12 = (x1-m[1]) %*% (x2-m[2]) /
+   sqrt( (x1-m[1]) %*% (x1-m[1]) * (x2-m[2]) %*% (x2-m[2]) )
> r12
          [,1]
[1,] 0.7649509

> cor(x1,x2)  #use built-in function
[1] 0.7649509
```

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

So the correlation is

$$r_{12} \quad = \quad 787.326 \,/\, (\, 21.303 \text{ x } 48.314 \,) = 0.7650 \qquad \text{(VII.5)}$$

as is shown above using R.

Consider now the closely related problem of finding the covariance matrix for all 4 variables involved. This is the matrix **C** with elements (i and j denote column variables and k denotes a city row)

$$C_{i\,j} \quad = \quad \frac{\Sigma\, (x_{ki} - m_i)\,(x_{kj} - m_j)}{n - 1} \qquad k=1,2,...,10 \qquad \text{(VII.6)}$$

where n is the number of data rows involved (for our crime example, n = 10) and the sum is over rows. In matrix notation (note: now using $\mathbf{x}_k$ as the k-th row of data!), eq.(VII.6) can be written using the *outer* product as

$$\mathbf{C} \quad = \quad \Sigma\, (\mathbf{x}_k - \mathbf{m})^t \text{ x } (\mathbf{x}_k - \mathbf{m}) \,/\, (n\text{-}1) \qquad k=1,2,..,10 \qquad \text{(VII.7)}$$

where the sum is over cities (data rows), and **m** is the vector of expected values for the $\mathbf{x}_k$.

This formula can be verified for our example using R:

```
> #find covariance using outer product formula
> cv = matrix(rep(0,16), ncol=4)  #set to 4x4 zero matrix
> for (k in 1:nrow(crime)) {
+   x = unlist(crime[k,]-colMeans(crime)) #convert to row vector format and
subtract mean
+   cv = cv + x %o% x   #accumulate outer products over data
+ }
> cv/(nrow(crime)-1)  #divide by degrees of freedom to get covariance
                Murder       Rape   Robbery AggrAssault
Murder        50.42489   87.48067  1315.022    961.2747
Rape          87.48067  259.36544  2457.990   2116.1737
Robbery     1315.02222 2457.99000 40760.064  28509.3856
AggrAssault  961.27467 2116.17367 28509.386  21696.1957

> cov(crime) #use built-in function to find covariance matrix
                Murder       Rape   Robbery AggrAssault
Murder        50.42489   87.48067  1315.022    961.2747
Rape          87.48067  259.36544  2457.990   2116.1737
Robbery     1315.02222 2457.99000 40760.064  28509.3856
AggrAssault  961.27467 2116.17367 28509.386  21696.1957
```

(The complicated code simply means that eq.(VII.7) gives the same covariance matrix that the built-in function in R does.)

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Finally, suppose we wish to predict the expected Murder rate knowing the other 3 variables, using the multiple linear regression model:

$$\text{Murder} = b_0 + b_1 \text{ Rape} + b_2 \text{ Robbery} + b_3 \text{ AggrAssault} \qquad \text{(VII.8)}$$

We need to find the best fit estimates for b_0, b_1, b_2, b_3 (assuming normally distributed errors).

Let **X** be the 10 x 4 matrix consisting of a column of 10 ones followed by the last 3 columns of data (Rape, Robbery and Aggravated Assault), and **y** be the column vector of data corresponding to Murder rate. Then eq.(VII.8) can be written in matrix form for the data as

$$\mathbf{y} = \mathbf{X\,b} + \mathbf{e} \qquad \text{(VII.9)}$$

where $\mathbf{b} = [b_0\ b_1\ b_2\ b_3]^t$ is the column vector of unknown parameters and **e** is a column vector of residual errors which is uncorrelated with the predictors **X** for the model on the data.

Multiplying both sides of eq.(VII.9) by $\mathbf{X}^t$, and using $\mathbf{X}^t\,\mathbf{e} = \mathbf{0}$ (i.e., the errors are uncorrelated with the predictors),

$$\mathbf{X}^t\,\mathbf{X\,b} = \mathbf{X}^t\,\mathbf{y} \qquad \text{(VII.10)}$$

This is the matrix version of the “normal equations” used to find the best estimates for the observed data of b_0, b_1, b_2, b_3.

To solve eq.(VII.10) for **b**, set

$$\mathbf{A} = \mathbf{X}^t\,\mathbf{X} \qquad \text{(VII.11)}$$

and multiply both sides of eq.(VII.10) by $\mathbf{A}^{-1}$:

$$\mathbf{A}^{-1}\,(\mathbf{A}\,b) = (\mathbf{X}^t\,\mathbf{X})^{-1}\,(\mathbf{X}^t\,\mathbf{X\,b}) = (\mathbf{X}^t\,\mathbf{X})^{-1}\,\mathbf{X}^t\,\mathbf{y} = \mathbf{A}^{-1}\,\mathbf{X}^t\,\mathbf{y} \qquad \text{(VII.12a)}$$

$$\mathbf{b} = \mathbf{A}^{-1}\,\mathbf{X}^t\,\mathbf{y} = (\mathbf{X}^t\,\mathbf{X})^{-1}\,\mathbf{X}^t\,\mathbf{y} \qquad \text{(VII.12b)}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Now let's see how eq.(VII.12b) calculates out for our crime data using R:

```
> X = as.matrix(cbind(Intercept=rep(1,10), crime[,2:4])) #make X have column of
1's and last 3 columns of crime
> X
             Intercept Rape Robbery AggrAssault
Dallas               1 53.3   553.9       584.2
Houston              1 41.2   548.3       561.7
Las Vegas            1 54.6   409.0       507.7
Los Angeles          1 27.3   370.0       377.2
New York             1 13.1   287.9       329.6
Philadelphia         1 65.5   749.1       720.1
Phoenix              1 36.2   287.5       398.5
San Antonio          1 39.8   179.6       388.7
San Diego            1 27.5   170.8       300.8
San Jose             1 23.6   111.9       248.2

> y = crime[,1] #set y = Murder
> y
 [1] 15.0 18.2 11.6 12.4  7.3 27.7 15.4  9.2  5.4  3.2

> A = t(X) %*% X  #form A matrix as XtX
> A
            Intercept      Rape   Robbery AggrAssault
Intercept        10.0    382.10    3668.0      4416.7
Rape            382.1  16934.33  162276.2    187807.7
Robbery        3668.0 162276.19 1712263.0   1876630.0
AggrAssault    4416.7 187807.67 1876630.0   2145989.6

> Ainv = chol2inv(chol(A))  #find inverse using special functions
> Ainv
            [,1]          [,2]          [,3]          [,4]
[1,]  6.08595788  0.1339017647  2.029806e-02 -0.0419944110
[2,]  0.13390176  0.0051965710  5.085056e-04 -0.0011750461
[3,]  0.02029806  0.0005085056  8.345017e-05 -0.0001592537
[4,] -0.04199441 -0.0011750461 -1.592537e-04  0.0003289949

> b = Ainv %*% t(X) %*% y #get coefficients
> b
            [,1]
[1,] -5.11756919
[2,] -0.05621700
[3,]  0.01023029
[4,]  0.03634649

> summary(lm(Murder ~ Rape + Robbery + AggrAssault, data=crime)) #do fit using
linear model

Residuals:
    Min      1Q  Median      3Q     Max
-3.7862 -1.4825 -0.4565  1.3153  5.1273

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.11757    7.96575  -0.642    0.544
Rape        -0.05622    0.23277  -0.242    0.817
Robbery      0.01023    0.02950   0.347    0.741
AggrAssault  0.03635    0.05857   0.621    0.558

Residual standard error: 3.229 on 6 degrees of freedom
Multiple R-squared: 0.8622,     Adjusted R-squared: 0.7932
F-statistic: 12.51 on 3 and 6 DF,  p-value: 0.005425
```

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

As can be seen, eq.(VII.12b) gives the same numerical estimates as the linear regression modeling built-in function of R.

None of the coefficients in the fit are statistically significant at the 0.05 level, although the entire model itself is significant (p-value = 0.005), suggesting perhaps a latent factor of “overall crime rate”.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

VIII. LINEAR DEPENDENCE AND INDEPENDENCE:

In doing an example of matrix inversion in Section VI we found the matrix was singular because we could construct one of the rows from a scalar combination of the other rows:

$$- \mathbf{A}_{1.} + 2\,\mathbf{A}_{2.} - \mathbf{A}_{3.} = 0 \qquad \text{(VIII.1)}$$

for the case show there. When such a “linear combination” of the rows of columns of a matrix **A** can be found which equals zero, as in eq.(VIII.1), then we will end up with a zero on the diagonal of the row echelon form when we try to invert **A**. This means the process must stop, and the inverse cannot be found. So such a matrix **A** is “non-invertible”, or “singular” (because of the *singular* problem we encountered, i.e., dividing by zero).

In general, for a set of vectors (row or column) $\mathbf{u}_1, ..., \mathbf{u}_n$, the vector $\mathbf{w}$ given by

$$\mathbf{w} \quad = \quad c_1\,\mathbf{u}_1 + c_2\,\mathbf{u}_2 + \ldots + c_n\,\mathbf{u}_n \qquad \text{(VIII.2)}$$

is called a “linear combination” of $\mathbf{u}_1, ..., \mathbf{u}_n$, where $c_1, ..., c_n$ are scalars. Note that eq.(VIII.2) could also be written

$$\mathbf{U}\,\mathbf{c} \quad = \quad \mathbf{w} \qquad \text{(VIII.3)}$$

where $\mathbf{u}_1, ..., \mathbf{u}_n$ are the column vectors of a matrix $\mathbf{U}$, and $\mathbf{c} = [c_1\ c_2\ ...\ c_n\ ]^t$.

If, for a given set of vectors $\mathbf{u}_1, ..., \mathbf{u}_n$ an equation of the form of eq.(VIII.2) can be found such that $\mathbf{w} = \mathbf{0}$, i.e.,

$$c_1\,\mathbf{u}_1 + c_2\,\mathbf{u}_2 + \ldots + c_n\,\mathbf{u}_n = \mathbf{0} \qquad \text{(VIII.4)}$$

then the set of vectors $\mathbf{u}_1, ..., \mathbf{u}_n$ are said to be “linearly dependent”. If no such set of $c_1, ..., c_n$ can be found such that eq.(VIII.4) holds (i.e., any such $\mathbf{w} \neq 0$), then $\mathbf{u}_1, ..., \mathbf{u}_n$ are said to be “linearly independent”. Alternatively, we can say that if

$$\mathbf{U}\,\mathbf{c} \quad = \mathbf{0} \qquad \text{(VIII.5)}$$

has any solution *other than* $\mathbf{c} = \mathbf{0}$, then $\mathbf{U}$ is singular, and $\mathbf{u}_1, ..., \mathbf{u}_n$ are *linearly dependent*. Contrariwise, if $\mathbf{U}$ is invertible, then eq.(VIII.5) has the *unique solution* $\mathbf{c} = \mathbf{0}$, and $\mathbf{u}_1, ..., \mathbf{u}_n$ are *linearly independent*.

In two dimensions (n = 2), if $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent, then they lie on a plane, but not on a line, so they are suitable for functioning as coordinate axes on the plane via eq.(VIII.2). In such a case, $\mathbf{u}_1$ and $\mathbf{u}_2$ are said to “span” the plane. If $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly dependent, however, they are collinear, and can’t supply two independent directions to serve as coordinate axes.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

In three dimensions (n = 3), if $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$ are linearly independent, they again can function as 3 coordinate axes pointing in 3 independent directions. If they are linearly dependent, then they must lie on a plane, or a line, or even a point (if $\mathbf{u}_1 = \mathbf{u}_2 = \mathbf{u}_3 = \mathbf{0}$), and not in 3 independent directions.

A similar situation exists for n > 3 with more complexity. In general, however, $\mathbf{u}_1, ..., \mathbf{u}_n$, do not span n independent dimensions unless they are linearly independent.

Now consider $\mathbf{u}_1 = \mathbf{A}_{1\,.}$, $\mathbf{u}_2 = \mathbf{A}_{2\,.}$, ..., $\mathbf{u}_n = \mathbf{A}_{n\,.}$, the row vectors of an n x n matrix **A**. One way to find out if $\mathbf{u}_1, ..., \mathbf{u}_n$ are linearly independent is to check if **A** has an inverse. If it does, then $\mathbf{u}_1, ..., \mathbf{u}_n$ are linearly independent. If not, not all of them are. The number of rows in the row echelon form that can be found before encountering an all-zero row is called the "rank" of the matrix **A**, and is the maximum number of rows of **A** (i.e., $\mathbf{u}_1, ..., \mathbf{u}_n$) that are still linearly independent. To put it another way, the rank(**A**) is the number of independent dimensions that the row vectors of **A** span. If rank(**A**) = n, then **A** is said to be of "full rank".

Row operations can reduce non-square matrices to row echelon form as well, using the same process.

A similar discussion and terminology applies to columns of **A** and its column echelon form.

It is not intuitively obvious, but it can be shown the column rank will always be equal to the row rank, so the term "rank" need only be used. For an n x m matrix **A**, rank(**A**) must obviously be no more than min(n, m).

In summary, if a square n x n matrix **A** is nonsingular, it is invertible, rank(**A**) = n, its row vectors span n dimensions and are linearly independent, and its column vectors span n dimensions and are linearly independent. Any one of the following conditions implies the others:

1. **A** is nonsingular.

2. $\mathbf{A}\,\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.

3. **A** is row equivalent (i.e., can be reduced to using elementary row operations) to the identity matrix.

4. $\mathbf{A}\,\mathbf{x} = \mathbf{b}$ has a unique solution for all **b**.

5. det(**A**) is nonzero. (See Section XIII for determinants.)

6. **A** has rank n, column rank n and row rank n.

7. The rows of **A** are linearly independent.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

8. The columns of **A** are linearly independent.

9. The rows of **A** span n dimensions.

10. The columns of **A** span n dimensions.

11. The “kernel”, “nullspace” or “nullity” of **A** (i.e., the solutions of $\mathbf{A}\ \mathbf{x} = \mathbf{0}$) is the unique vector $\mathbf{x} = \mathbf{0}$.

EXERCISES

In each case, find a row echelon form and the rank of the matrix given.

Q1. $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ A1. $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ rank = 1

Q2. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ & & \end{bmatrix}$ A2. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ & & \end{bmatrix}$ rank = 1

Q3. $\begin{bmatrix} 1 & 0 & \\ 0 & 0 & \\ 0 & 0 & \end{bmatrix}$ A3. $\begin{bmatrix} 1 & 0 & \\ 0 & 0 & \\ 0 & 0 & \end{bmatrix}$ rank = 1

Q4. $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ A4. $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ rank = 1

Q5. $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ & & \end{bmatrix}$ A5. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ & & \end{bmatrix}$ rank = 1

Q6. $\begin{bmatrix} 1 & 0 & \\ 1 & 0 & \\ 0 & 0 & \end{bmatrix}$ A6. $\begin{bmatrix} 1 & 0 & \\ 0 & 0 & \\ 0 & 0 & \end{bmatrix}$ rank = 1

Q7. $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ A7. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ rank = 2

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Q8. $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$ A8. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ rank = 2

Q9. $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ A9. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ rank = 2

Q10. $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ A10. $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ rank = 2

Q11. $\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$ A11. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ rank = 2

Q12. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ A12. $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ rank = 1

Q13. $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$ A13. $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ rank = 1

Q14. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$ A14. $\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ rank = 1

Q15. $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ A15. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ rank = 2

Q16. $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$ A16. $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ rank = 2

Q17. $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix}$ A17. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ rank = 2

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Q18. $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ A18. $\begin{bmatrix} 1 & 1/2 \\ 0 & 1 \end{bmatrix}$ rank = 2

Q19. $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ & & \end{bmatrix}$ A19. $\begin{bmatrix} 1 & 1/2 & 1/2 \\ 0 & 1 & 1/3 \\ & & \end{bmatrix}$ rank = 2

Q20. $\begin{bmatrix} 2 & 1 & \\ 1 & 2 & \\ 1 & 1 & \end{bmatrix}$ A20. $\begin{bmatrix} 1 & 1/2 & \\ 0 & 1 & \\ 0 & 0 & \end{bmatrix}$ rank = 2

Q21. $\begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$ A21. $\begin{bmatrix} 1 & 2/3 & 1/3 \\ 0 & 1 & 1/5 \\ 0 & 0 & 1 \end{bmatrix}$ rank = 3

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

IX. SOLUTION OF SYSTEMS OF LINEAR EQUATIONS:

Consider the single (scalar) equation

$$a\,x = b \tag{IX.1}$$

for a and b constants, and x unknown. This equation has an unique solution if $a \neq 0$, and no solution at all if $a = 0$.

Now consider the 2-dimensional system of equations

$$\mathbf{A\,x} = \begin{bmatrix} a11 & a12 \\ a21 & a22 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \end{bmatrix} = \begin{bmatrix} b1 \\ b2 \end{bmatrix} = \mathbf{b} \tag{IX.2}$$

for **A** and **b** constant, and **x** unknown.

If **A** = 0, i.e., rank(**A**) = 0, then there are no solutions. If **A** is invertible ($\det(A) \neq 0$), i.e., rank(**A**) = 2, there is a unique solution given by

$$\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b} \tag{IX.3}$$

If rank(**A**) = 1, however, then one row of **A** is a multiple of the other. If the elements of **b** are correspondingly multiples of each other, then the equations are said to be "consistent", and there are an infinite number of solutions parameterized by the 2 – rank(**A**) = 1 degrees of freedom present. If the elements of **b** are not in correspondence with the relation of the rows of **A**, then the equations are said to be "inconsistent", and there are no solutions.

As a simple example of this, suppose

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{IX.4}$$

Obviously the rows of **A** are the same (i.e., the multiple is 1). If $\mathbf{b} = [2\ 2]^t$, for example, the two equations are equivalent ("consistent") as

$$x_1 + x_2 = 2 \tag{IX.5}$$

with solutions

$$x_1 = 2 - x_2 \qquad \text{for any } x_2 \tag{IX.6}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

or, in parametric form with parameter θ,

$$x_1 = 2 - \theta$$
$$x_2 = \theta \qquad \text{(IX.7)}$$

Now suppose that **b** = [1 2]^t, then the first equation becomes

$$x_1 + x_2 = 1 \qquad \text{(IX.8)}$$

but the second is

$$x_1 + x_2 = 2 \qquad \text{(IX.9)}$$

These two equations are obviously inconsistent with each other, so the system of equations **A x** = **b** are "inconsistent", and there are no solutions at all.

For n > 2 equations, the situation is similar, but more complex. If rank(**A**) = 0, there are no solutions. If rank(**A**) = n, there is a unique solution, given by eq.(IX.3). If 0 < rank(**A**) < n, there may be no solution if the system is inconsistent, or an infinite number of solutions indexed by n – rank(**A**) parameters.

A method for finding out if a system of equations **A x** = **b** are consistent is as follows:

1. Consider the "augmented matrix" **C** which is the n x (n+1) matrix formed by the columns of **A** with the **b** column attached at the right.

2. Compute rank(**A**) and rank(**C**) by reducing each to row echelon form and counting the non-zero (i.e., linearly independent) rows.

3. If rank(**C**) > rank(**A**), then the equations are inconsistent, and there are no solutions.

4. If rank(**C**) = rank(**A**), then the equations are consistent, and there is a unique solution if rank(**A**) = n, and an infinite number if 0 < rank(**A**) < n.

All of these cases can be combined into a common method for finding a solution to the system of equations

$$\mathbf{A\,x} = \mathbf{b} \qquad (\text{given } \mathbf{A} \neq \mathbf{0}) \qquad \text{(IX.10)}$$

First, form the augmented matrix **C**, which is the coefficient matrix **A** with the column vector b appended.

Second, use elementary row operations to reduce **C** to row echelon form.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

If none of the first n elements of any row are all zeroes, rank(**A**) = n, and there is a unique solution found by back-substitution. Alternatively, you can further row reduce the first columns of **C** to a rank(**A**) x rank(**A**) identity matrix, in which case the (n+1)-th column will contain the unique solution.

If one or more rows of **C** have the first n elements of any row all zero, but the (n+1)-th element is non-zero, then the system of equations is inconsistent, and there is no solution.

In all other cases, there are an infinite number of solutions, with n – rank(**A**) unknowns indeterminable and arbitrary.

For an example of an inconsistent set of equations, consider

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 3 & 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \qquad \text{(IX.11)}$$

The augmented matrix **C** is then

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & 1 & -1 \\ 3 & 1 & 2 & 1 \end{bmatrix} \qquad \text{(IX.12)}$$

Row-reducing **C** to echelon form results in

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \text{(IX.13)}$$

The last row of eq.(IX.13) shows the system of equations is inconsistent (i.e., impossible to solve).

For an example of a consistent system of equation, consider

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 3 & 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \qquad \text{(IX.14)}$$

for which the augmented matrix **C** is

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & 1 & -1 \\ 3 & 1 & 2 & 0 \end{bmatrix} \qquad \text{(IX.15)}$$

Row-reducing **C** to echelon form results in

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{(IX.16)}$$

The last row of eq.(IX.16) shows the system of equations is consistent, although of rank 2.

Carrying out a further row operation to clear the row 1 column 2 element,

$$\begin{bmatrix} 1 & 0 & 1/2 & -1/2 \\ 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{(IX.17)}$$

which corresponds to

$$\begin{aligned} x_1 &= -1/2 - x_3/2 \\ x_2 &= 3/2 - x_3/2 \\ x_3 &= x_3 \end{aligned} \qquad \text{(IX.18)}$$

or, introducing the parameter θ for x_3 on the right-hand side of the equations,

$$\begin{aligned} x_1 &= -1/2 - \theta/2 \\ x_2 &= 3/2 - \theta/2 \\ x_3 &= \theta \end{aligned} \qquad \text{(IX.19)}$$

For example, for $\theta = 0$, $x_1 = -1/2$, $x_2 = 3/2$ and $x_3 = 0$. Similarly, for $\theta = 1$, $x_1 = -1$, $x_2 = 1$ and $x_3 = 1$. (You should verify that these are both solutions of **A x** = **b** with **A** and **b** given by eq.(IX.14).)

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

X. THE GENERALIZED INVERSE:

The ideas of the previous section can be carried farther by attempting to find something that functions similar to the inverse of a nonsingular square matrix, but can be found for rectangular and singular (i.e., not full rank) matrices as well.

We know the inverse $\mathbf{A}^{-1}$ of a square nonsingular matrix **A** satisfies, among other things,

$$\mathbf{A}\,\mathbf{A}^{-1} = \mathbf{I} \qquad \text{(X.1)}$$

This means that $\mathbf{A}\,\mathbf{A}^{-1}$ functions like **I**. In particular,

$$(\mathbf{A}\,\mathbf{A}^{-1})\,\mathbf{A} = \mathbf{A}\,(\mathbf{A}^{-1}\,\mathbf{A}) = \mathbf{A}\,\mathbf{A}^{-1}\,\mathbf{A} = \mathbf{A} \qquad \text{(X.2)}$$

This motivates a search for a matrix **G** such that

$$\mathbf{A}\,\mathbf{G}\,\mathbf{A} = \mathbf{A} \qquad \text{(X.3)}$$

where **A** can now be rectangular or square, or singular or nonsingular. If **A** is of order m x n then **G** is of order n x m. Eq.(X.3) is called the "First Penrose Condition" after the person who first developed the theory.

Any matrix **G** for which eq.(X.3) holds for a particular matrix **A** is called a "generalized inverse" of **A**. Another term used for **G** is "pseudoinverse". The matrix **G** is not uniquely determined by eq.(X.3), which is insufficient for this purpose. In fact, if **G** is a generalized inverse, then so are all matrices **H** which satisfy

$$\mathbf{H} = \mathbf{G}\,\mathbf{A}\,\mathbf{G} + (\mathbf{I} - \mathbf{G}\,\mathbf{A})\,\mathbf{T} + \mathbf{S}\,(\mathbf{I} - \mathbf{A}\,\mathbf{G}) \qquad \text{(X.4)}$$

for some matrices **T** and **S** of the right size. In fact, all generalized inverses of A can be found by using eq.(X.4) for all possible **T** and **S**, starting with one generalized inverse **G**.

Sometimes additional conditions (together with eq.(X.3) called the "Moore-Penrose Conditions") are added to uniquely specify a particular generalized inverse (the "Moore-Penrose Generalized Inverse") with additional desirable properties. These conditions are:

$$\mathbf{G}\,\mathbf{A}\,\mathbf{G} = \mathbf{G} \qquad \text{(X.5a)}$$

$$\mathbf{A}\,\mathbf{G} \text{ is symmetric} \qquad \text{(X.5b)}$$

$$\mathbf{G}\,\mathbf{A} \text{ is symmetric} \qquad \text{(X.5d)}$$

Note that if A is square and nonsingular, there is a unique solution to eq.(X.3), and this is $\mathbf{G} = \mathbf{A}^{-1}$.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Returning to eq.(X.3), there are several simple methods by which at least one **G** can be found.

First, if row operations, whose effective product is a matrix **R**, reduces **A** to the row echelon form

$$\mathbf{R}\,\mathbf{A} \quad = \quad \begin{bmatrix} T & U \\ 0 & 0 \end{bmatrix} \qquad \text{(X.6)}$$

where **T** is upper triangular and nonsingular and **U** is some other matrix, then

$$\mathbf{G} \quad = \quad \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \mathbf{R} \qquad \text{(X.7)}$$

is a generalized inverse of **A**, where $\mathbf{S} = \mathbf{T}^{-1}$.

Second, if the upper rank(**A**) x rank(**A**) block $\mathbf{A}_{11}$ of A is nonsingular, then

$$\mathbf{G} \quad = \quad \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{(X.8)}$$

is a generalized inverse of **A**, where $\mathbf{W} = \mathbf{A}_{11}^{-1}$.

There are another of other special cases, which won't be described here.

For a general matrix **A**, Hence's algorithm for finding a generalized inverse **G** is:

1. Find in **A** a nonsingular submatrix of order rank(**A**). Denote it by **W**.
2. Invert and transpose **W**, i.e., $(\mathbf{W}^{-1})^t$. *Note the order of operations!*
3. In **A**, replace each element of **W** by the corresponding element of $(\mathbf{W}^{-1})^t$.
4. Replace all other elements of **A** by 0.
5. Transpose the resulting matrix.
6. The result is **G**, a generalized inverse of **A**.

A common notation for **G** is $\mathbf{A}^-$, which will be used from now on.

Also, again note that if **A** is of order m x n then **G** is of order n x m.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

EXAMPLE: FINDING A GENERALIZED INVERSE

Consider the matrix **A** given by

$$\mathbf{A} \quad = \quad \begin{bmatrix} 2 & 3 & 1 & -1 \\ 5 & 8 & 0 & 1 \\ 1 & 2 & -2 & 3 \\ \end{bmatrix} \qquad \text{(X.9)}$$

First we need to find the rank(**A**). Clearly it must be 3 or less. Using row operations and reducing **A** to row echelon form, or simply noting that row 2 is the sum of 2 times the first row plus the third row, and rows 1 and 3 are linearly independent (not multiples), we find the rank(**A**) = 2.

Let the needed submatrix **W** be the upper left 2x2 submatrix:

$$\mathbf{W} \quad = \quad \begin{bmatrix} 2 & 3 \\ 5 & 8 \end{bmatrix} \qquad \text{(X.10)}$$

The determinant of **W** is 16 – 15 = 1, so the inverse of **W** is

$$\mathbf{W}^{-1} \quad = \quad \begin{bmatrix} 8 & -3 \\ -5 & 2 \end{bmatrix} \qquad \text{(X.11)}$$

with a transpose of

$$(\mathbf{W}^{-1})^{t} \quad = \quad \begin{bmatrix} 8 & -5 \\ -3 & 2 \end{bmatrix} \qquad \text{(X.12)}$$

Now we substitute for W in A, zeroing the 3rd and 4th columns and the 3rd row:

$$\begin{bmatrix} 8 & -5 & 0 & 0 \\ -3 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \end{bmatrix} \qquad \text{(X.13)}$$

Finally, we transpose this result to get a generalized inverse $\mathbf{A}^{-}$:

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

$$\mathbf{A}^{-} \quad = \quad \begin{bmatrix} 8 & -3 & 0 \\ -5 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{(X.14)}$$

$\mathbf{A}^{-}$ is verified as a generalized inverse by computing $\mathbf{A}\ \mathbf{A}^{-}\ \mathbf{A}$ and showing it is equal to $\mathbf{A}$.

Using R:

```
> A<- matrix(c(2,3,1,-1, 5,8,0,1, 1,2,-2,3), ncol=4, byrow=TRUE)
> A
     [,1] [,2] [,3] [,4]
[1,]    2    3    1   -1
[2,]    5    8    0    1
[3,]    1    2   -2    3

> G<- matrix(c(8,-3,0, -5,2,0, 0,0,0, 0,0,0), ncol=3, byrow=TRUE)
> G
     [,1] [,2] [,3]
[1,]    8   -3    0
[2,]   -5    2    0
[3,]    0    0    0
[4,]    0    0    0

> A %*% G %*% A
     [,1] [,2] [,3] [,4]
[1,]    2    3    1   -1
[2,]    5    8    0    1
[3,]    1    2   -2    3
```

The final result is indeed identical to the original matrix **A**.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

XI. SOLUTION OF EQUATIONS USING THE GENERALIZED INVERSE:

There is another approach to the generalized inverse $\mathbf{A}^-$ of $\mathbf{A}$ that is closer to the way that it is used. The solution to a (consistent) system of equations

$$\mathbf{A}\,\mathbf{x} = \mathbf{b} \qquad \text{(XI.1)}$$

is given by

$$\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b} \qquad \text{(XI.2)}$$

when $\mathbf{A}$ is square and nonsingular. By analogy, we would like the "generalized" inverse $\mathbf{A}^-$ of $\mathbf{A}$ to satisfy

$$\mathbf{x} = \mathbf{A}^-\,\mathbf{b} + \mathbf{F}\,\boldsymbol{\theta} \qquad \text{(XI.3)}$$

where $\mathbf{F}$ is a matrix of appropriate order and $\boldsymbol{\theta}$ is a vector of variable parameters of order the same as $\mathbf{x}$ or the number of columns of $\mathbf{A}$. For $\mathbf{A}$ nonsingular, $\mathbf{F} = \mathbf{0}$, and eq.(XI.3) becomes eq.(XI.2).

In eq.(XI.3), it can be shown that $\mathbf{F} = (\mathbf{A}^-\,\mathbf{A} - \mathbf{I})$, or

$$\mathbf{x} = \mathbf{A}^-\,\mathbf{b} + (\mathbf{A}^-\,\mathbf{A} - \mathbf{I})\,\boldsymbol{\theta} \qquad \text{(XI.4)}$$

If both sides of eq.(XI.4) are multiplied by $\mathbf{A}$,

$$\mathbf{A}\,\mathbf{x} = \mathbf{A}\,\mathbf{A}^-\,\mathbf{b} + (\mathbf{A}\,\mathbf{A}\text{-}\,\mathbf{A} - \mathbf{A})\,\boldsymbol{\theta} \qquad \text{(XI.5a)}$$

$$= \mathbf{A}\,\mathbf{A}^-\,\mathbf{b} + (\mathbf{A} - \mathbf{A})\,\boldsymbol{\theta} \qquad \text{(XI.5b)}$$

$$= \mathbf{A}\,\mathbf{A}^-\,\mathbf{b} \qquad \text{(XI.5c)}$$

But, because

$$\mathbf{A}\,\mathbf{A}^-\,\mathbf{A} = \mathbf{A} \qquad \text{(XI.6)}$$

then

$$\mathbf{A}\,\mathbf{A}^-\,\mathbf{A}\,\mathbf{x} = \mathbf{A}\,\mathbf{x} \qquad \text{(XI.7a)}$$

$$\mathbf{A}\,\mathbf{A}^-\,(\mathbf{A}\,\mathbf{x}) = (\mathbf{A}\,\mathbf{x}) \qquad \text{(XI.7b)}$$

$$\mathbf{A}\,\mathbf{A}^-\,\mathbf{b} = \mathbf{b} \qquad \text{(XI.7c)}$$

So eq. (XI.5c) may be written

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

$$\mathbf{A}\,\mathbf{x} = \mathbf{b} \qquad \text{(XI.8)}$$

which proves eq.(XI.4), as desired. Note that, because **θ** is composed of arbitrary parameters, eq.(XI.4) could also have been written without loss of generality as

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\,\boldsymbol{\theta}' \qquad \text{(XI.9)}$$

with **θ**’ = **-** **θ**.

Eq.(XI.4) or eq.(XI.9) gives the general solution to eq.(XI.1) for any values of the arbitrary **θ** (or **θ**’).

As an example of finding the general solution to eq.(XI.1), consider

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 & 3 \\ 1 & 1 & 1 & 2 \\ 3 & 5 & 1 & 4 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \text{(XI.10)}$$

First, we need to find the rank of **A** and whether the equations are consistent. Forming the augmented matrix by adding **b** as a column to **A** and then row reducing:

$$\begin{bmatrix} 2 & 3 & 1 & 3 & 1 \\ 1 & 1 & 1 & 2 & 1 \\ 3 & 5 & 1 & 4 & 1 \end{bmatrix} \text{->} \begin{bmatrix} 1 & 3/2 & 1/2 & 3/2 & 1/2 \\ 1 & 1 & 1 & 2 & 1 \\ 3 & 5 & 1 & 4 & 1 \end{bmatrix} \text{->} \qquad \text{(XI.11a)}$$

$$\begin{bmatrix} 1 & 3/2 & 1/2 & 3/2 & 1/2 \\ 0 & -1/2 & 1/2 & 1/2 & 1/2 \\ 0 & 1/2 & -1/2 & -1/2 & -1/2 \end{bmatrix} \text{->} \qquad \text{(XI.11b)}$$

$$\begin{bmatrix} 1 & 3/2 & 1/2 & 3/2 & 1/2 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 1/2 & -1/2 & -1/2 & -1/2 \end{bmatrix} \text{->} \qquad \text{(XI.11c)}$$

$$\begin{bmatrix} 1 & 3/2 & 1/2 & 3/2 & 1/2 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{(XI.11d)}$$

Because the entire last row is entirely zero, the equations are consistent, and rank(**A**) = 2.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Using the algorithm at the end of Section X to find a generalized inverse $\mathbf{A}^-$ of $\mathbf{A}$, we choose a 2 x 2 submatrix $\mathbf{W}$ of $\mathbf{A}$ which is invertible. We choose the upper-left 2x2 submatrix:

$$\mathbf{W} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} \tag{XI.12}$$

which has inverse $\mathbf{W}^{-1}$ and its transpose $(\mathbf{W}^{-1})^t$ given by

$$\mathbf{W}^{-1} = \begin{bmatrix} 1 & -3 \\ -1 & 2 \end{bmatrix} / [(2)(1)-(3)(1)] = \begin{bmatrix} -1 & 3 \\ 1 & -2 \end{bmatrix} \tag{XI.13a}$$

$$(\mathbf{W}^{-1})^t = \begin{bmatrix} -1 & 1 \\ 3 & -2 \end{bmatrix} \tag{XI.13b}$$

Now we construct the generalized inverse $\mathbf{A}^-$ by replacing the submatrix in $\mathbf{A}$ by the values in eq.(XI.13b) and filling the rest with zeroes, transposing, and noting that $\mathbf{A}^-$ is of order 4 x 3:

$$\mathbf{A}^- = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 3 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ & & & \end{bmatrix}^t \tag{XI.14a}$$

$$= \begin{bmatrix} -1 & 3 & 0 & \\ 1 & -2 & 0 & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \end{bmatrix} \tag{XI.14b}$$

Computing the intermediate result $\mathbf{C} = (\mathbf{A}^- \mathbf{A} - \mathbf{I})$,

$$\mathbf{C} = \begin{bmatrix} -1 & 3 & 0 & \\ 1 & -2 & 0 & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 & 3 \\ 1 & 1 & 1 & 2 \\ 3 & 5 & 1 & 4 \\ & & & \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{XI.15a}$$

$$= \begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{XI.15b}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 2 & 3 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{XI.15c}$$

The general solution to eq.(XI.1) for this case is then

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{b} + \mathbf{C}\,\boldsymbol{\theta} \tag{XI.16a}$$

$$= \begin{bmatrix} -1 & 3 & 0 & \\ 1 & -2 & 0 & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 2 & 3 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix} \tag{XI.16b}$$

$$= \begin{bmatrix} 2 \\ -1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 2r+3s \\ -r-s \\ -r \\ -s \end{bmatrix} \tag{XI.16c}$$

$$= \begin{bmatrix} 2+2r+3s \\ -1-r-s \\ -r \\ -s \end{bmatrix} \tag{XI.16d}$$

for any values of r and s. Note also that the constant vector in eq.(XI.16c) is, indeed, a solution of the system eq.(XI.1):

$$\mathbf{A}\,\mathbf{x} = \begin{bmatrix} 2 & 3 & 1 & 3 \\ 1 & 1 & 1 & 2 \\ 3 & 5 & 1 & 4 \\ & & & \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \; \end{bmatrix} = \mathbf{b} \tag{XI.17}$$

The same can be shown for the general solution in eq.(XI.16d).

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

APPENDIX: FURTHER COMMENTS ON CONSISTENT SYSTEMS OF LINEAR EQUATIONS

A system of linear equations

$$\mathbf{A\,x} = \mathbf{b} \qquad \text{(XI.18)}$$

is said to be "consistent" if any linear relations between rows of **A** also exist (are satisfied) by the corresponding elements of **b**. I.e., the relations are consistent between A and **b**.

When the matrix **A** has full row rank, there are no linear relations between its rows, because they are all linearly independent. So if **A** has full row rank, eq.(XI.18) must be consistent.

The set of linear equations eq.(XI.18) can be solved if and only if the system is consistent. Inconsistent equations have zero solutions.

In general, eq.(XI.18) will have zero, one or an infinite number of solutions. There are zero solutions if eq.(XI.18) is inconsistent between **A** and **b**. If **A** and **b** are consistent, there will be one solution if **A** has a rank equal to the order of **x** (i.e., **A** has full column rank), so that there is one independent equation to "pin down" the value of each variable in **x**, or an infinite number of solutions if rank(**A**) < order(**x**).

You test for consistency by forming the "augmented" matrix by appending the column **b** to those of **A**, and then find the rank of the augmented matrix by the usual row reduction to echelon form. Eq.(XI.18) is consistent if and only if the rank of the augmented matrix is equal to the rank of **A**. I.e., **b** does not add anything "new" to the columns of **A**.

For a real matrix **A**, the product matrix $\mathbf{A}^t\,\mathbf{A}$ is non-negative definite (it is the sum of squares of elements of **A**), and is positive definite when **A** has full column rank. If **A** has full column rank, then $(\mathbf{A}^t\,\mathbf{A})^{-1}\,\mathbf{A}^t$ is a left inverse of **A**, and if **A** has full row rank then $\mathbf{A}^t\,(\mathbf{A}\,\mathbf{A}^t)^{-1}$ is a right inverse of **A**. I.e.,

$$(\mathbf{A}^t\,\mathbf{A})^{-1}\,\mathbf{A}^t\,\mathbf{A} = \mathbf{I} \qquad \text{(XI.19a)}$$

$$\mathbf{A}\,\mathbf{A}^t\,(\mathbf{A}\,\mathbf{A}^t)^{-1} = \mathbf{I} \qquad \text{(XI.19b)}$$

(Eqs.(XI.19) should both be obvious by inspection.)

Suppose eq.(XI.18 is to be solved, **A** has full column rank and **A** and **b** are consistent. There is then a unique solution, which is given by

$$\mathbf{x} = (\mathbf{A}^t\,\mathbf{A})^{-1}\,\mathbf{A}^t\,\mathbf{b} \qquad \text{(XI.20)}$$

which follows by multiplying both sides of eq.(XI.18) by $(\mathbf{A}^t\,\mathbf{A})^{-1}\,\mathbf{A}^t$ on the left.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

When eq.(XI.18) is consistent, but **A** does not have full column rank, there are an infinite number of solutions. These are given by the generalized inverse solution

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{b} + (\mathbf{A}^{-}\mathbf{A} - \mathbf{I})\,\boldsymbol{\theta} \qquad \text{(XI.21)}$$

where $\mathbf{A}^{-}$ is a generalized inverse of **A**.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

XII. HOMOGENEOUS SYSTEMS OF EQUATIONS:

We now consider the special case where the right-hand side vector **b** = **0**, or

$$\mathbf{A}\,\mathbf{x} \quad = \quad \mathbf{0} \qquad \text{(XII.1)}$$

The set of all **x** vectors which solve eq.(XII.1) are called the “nullspace”, “nullity” or “kernel” of **A**, and eq.(XII.1) is called a “homogeneous” system of linear equations.

It should be obvious that **x** = **0** will always be one of the solutions of eq.(XII.1), so it is called the “trivial” solution. It should also be obvious that a homogeneous system of equations are never inconsistent (hint: the right-hand side column vector is all zero).

If **A** is square and nonsingular, or, more generally, if rank(**A**) = order(**x**), there is a unique solution to eq.(XII.1), and that is the trivial solution **x** = **0**. If this case, the number of free equations (rows) equals the number of free parameters (elements of **x**), and the solution space has zero degrees of freedom, or a point.

In the case where rank(**A**) < order(**x**), there are fewer free equations (rows) in **A** than there are free parameters in **x**, so there are d = order(**x**) – rank(**A**) degrees of freedom left free as parameters to index the solution.

As an example, consider the matrix **A** used in Section X:

$$\mathbf{A} \quad = \quad \begin{bmatrix} 2 & 3 & 1 & 3 \\ 1 & 1 & 1 & 2 \\ 3 & 5 & 1 & 4 \\ & & & \end{bmatrix} \qquad \text{(XII.2)}$$

This matrix has rank = 2, as was determined in Section X, but **x** is of order 4. So there are d = 4 – 2 = 2 degrees of freedom left to characterize the nullspace of **A**.

To find a solution to eq.(XII.1) for this A, we could row reduce the augmented matrix as far as possible towards a 2 x 2 identity matrix, and then parameterize the resulting equations. But having obtained the general solution via the generalized inverse in eq.(XI.16a), it is easier to just substitute the known $\mathbf{A}^-$ and **C** matrices for **b** = **0**:

$$\mathbf{x} \quad = \quad \mathbf{A}^-\mathbf{b} \; + \mathbf{C}\,\boldsymbol{\theta} \qquad \text{(XII.3a)}$$

$$= \quad \begin{bmatrix} -1 & 3 & 0 & \\ 1 & -2 & 0 & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 2r+3s \\ -r-s \\ -r \\ -s \end{bmatrix} \qquad \text{(XII.3b)}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

$$= \quad \mathbf{0} \quad + \quad \begin{bmatrix} 2r+3s \\ -r-s \\ -r \\ -s \end{bmatrix} \qquad \text{(XII.3c)}$$

$$= \quad \begin{bmatrix} 2r+3s \\ -r-s \\ -r \\ -s \end{bmatrix} \qquad \text{(XII.3d)}$$

Note how the first term in eq.(XII.3a) will always be 0 for b = 0, so the general solution to the homogeneous eq.(XII.1) is given by the second term **C θ**. Note also that eq.(XII.3d) involves two free parameters, denoted r and s, as expected for d = 2.

Eq.(XII.3d) can be rewritten

$$\mathbf{x} \quad = \quad \begin{bmatrix} 2r+3s \\ -r-s \\ -r \\ -s \end{bmatrix} \qquad \text{(XII.4a)}$$

$$= \quad r \begin{bmatrix} 2 \\ -1 \\ -1 \\ 0 \end{bmatrix} \quad + \quad s \begin{bmatrix} 3 \\ -1 \\ 0 \\ -1 \end{bmatrix} \qquad \text{(XII.4b)}$$

which can be seen to be a linear combination with coefficients r and s of two linearly independent constant vectors, each of which solve eq.(XII.1). So the null space of **A** is spanned by the two vectors shown in eq.(XII.4b).

For the "inhomogeneous" equations case eq.(XI.1) with $\mathbf{b} \neq \mathbf{0}$, the first term in eq.(XI.4) or eq.(XI.16a) is a constant vector which solves eq.(XI.1), and is called a "particular" solution to eq.(XI.1), denoted $\mathbf{x}_p$. The second term in eq.(XI.4) or eq.(XI.16a) is the "general" solution to the homogenous system of equations eq.(XII.1), and is of the form of eq.(XII.4b), a linear combination of d = order(**x**) – rank(**A**) different linearly independent solutions $\mathbf{x}_1 \ldots \mathbf{x}_d$ of eq.(XII.1). So the general solution of eq.(XII.1) can be written

$$\mathbf{x} \quad = \quad \mathbf{x}_p \quad + \quad c_1\,\mathbf{x}_1 + c_2\,\mathbf{x}_2 + \ldots + c_d\,\mathbf{x}_d \qquad \text{(XII.5)}$$

where $c_1 \ldots c_d$ are d scalar parameters indexing the nullspace solutions. Solutions of the form of eq.(XII.5) are characteristic of those of linear operator equations.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Note that, if d = 1, then eq.(XI.1) has solutions

$$\mathbf{x} = \mathbf{x}_p + c_1 \mathbf{x}_1 \tag{XII.6}$$

and eq.(XII.1) has solutions

$$\mathbf{x} = c_1 \mathbf{x}_1 \tag{XII.7}$$

so, if $\mathbf{x}_1$ is a solution of eq.(XII.1), then so is any multiple of it. This is generally not true if d > 1.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

XIII. DETERMINANT OF A SQUARE MATRIX:

The "determinant" is defined only for square matrices, and is a scalar calculated from products of the elements of the matrix with +1 or –1 coefficients determined by certain rules. The determinant of a matrix **A** is denoted |**A**| or det(**A**), with the former notation suggestive of a norm, although the determinant may be either positive or negative or zero. Most people first learn about determinants in secondary school algebra courses where linear equations are solved using Cramer's rule.

For a 1 x 1 matrix **A**, the determinant is simply the value of the single element a_{11}.

For a 2 x 2 matrix

$$\mathbf{A} = \begin{bmatrix} a11 & a12 \\ a21 & a22 \end{bmatrix} \qquad \text{(XIII.1)}$$

then

$$|\mathbf{A}| = a_{11}\, a_{22} - a_{12}\, a_{21} \qquad \text{(XIII.2)}$$

For example, if

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} \qquad \text{(XIII.3)}$$

then

$$|\mathbf{A}| = (2)(1) - (3)(1) = 2 - 3 = -1 \qquad \text{(XIII.4)}$$

For a 3 x 3 matrix

$$\mathbf{A} = \begin{bmatrix} a11 & a12 & a13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{bmatrix} \qquad \text{(XIII.5)}$$

then

$$|\mathbf{A}| = a_{11}\, a_{22}\, a_{33} - a_{11}\, a_{23}\, a_{32} - a_{12}\, a_{21}\, a_{33} + a_{12}\, a_{23}\, a_{31} + a_{13}\, a_{21}\, a_{32} - a_{13}\, a_{22}\, a_{31} \qquad \text{(XIII.6)}$$

Note that each term in eq.(XIII.6) is of the form $a_{1\,j1}\, a_{2\,j2}\, a_{3\,j3}$, where $(j_1\, j_2\, j_3)$ has the same elements as (1 2 3), but with a different order. The sign (+1 or –1) in front of each term in eq.(XIII.6) is +1 if $\{j_1, j_2, j_3\}$ is an even permutation of (1 2 3) and –1 if it is an odd permutation.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

A permutation is "even" or has sign +1 if an even number of transpositions will generate the permutation, and "odd" or has sign –1 if an odd number of transpositions are needed.

For example, the permutation (1 2 3) requires 0 transpositions, and so is even or of sign +1, which is the coefficient of the term $a_{11}\ a_{22}\ a_{33}$ in eq.(XIII.6). Contrariwise, the permutation (1 3 2) requires one transposition (of 2 and 3), and so is odd and of sign –1, as is the term $a_{11}\ a_{23}\ a_{32}$ in eq.(XIII.6). Similarly, the permutation (2 3 1) can be generated with two transpositions (1 and 2, followed by 1 and 3), and is even, so $a_{12}\ a_{23}\ a_{31}$ has a +1 coefficient in eq.(XIII.6).

Generally, for an n x n matrix **A**, the determinant is given by

$$| \mathbf{A} | \quad = \quad \sum \operatorname{sgn}(j_1\ j_2\ \ldots\ j_n)\ a_{1\,j1}\ a_{2\,j2}\ \ldots\ a_{n\,jn} \qquad \text{(XIII.7)}$$

where the sum is over all possible permutations $(j_1\ j_2\ \ldots\ j_n)$ of (1 2 ... n).

As an example, consider

$$\mathbf{A} \quad = \quad \begin{bmatrix} 2 & 3 & 1 \\ 1 & 1 & 1 \\ 3 & 5 & 1 \end{bmatrix} \qquad \text{(XIII.8)}$$

The determinant is given by

$$| \mathbf{A} | \quad = \quad (2)(1)(1) - (2)(1)(5) + (3)(1)(3) - (3)(1)(1) + (1)(1)(5) - (1)(1)(3) \qquad \text{(XIII.9a)}$$

$$= \quad +2 - 10 + 9 - 3 + 5 - 3 \qquad \text{(XIII.9b)}$$

$$= \quad 0$$

If the | **A** | = 0, as here, the matrix **A** is singular, noninvertible, and rank(**A**) < n.

Eq.(XIII.7) is not the only algorithm available to find | **A** |. Others include multiplying the diagonal elements after reduction to row echelon form, and a recursive method learned by most people in secondary school.

This recursive method is based upon expansion in "minors", where the "minor" corresponding to row i and column j is denoted $| \mathbf{A}_{i\,j} |$, and is defined as the determinant of the submatrix obtained by deleting the i-th row and j-th column of **A**. For example, for **A** given by eq.(XIII.8), the minor $| \mathbf{A}_{12} |$ is given by

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

$$| \mathbf{A}_{12} | = \begin{vmatrix} 1 & 1 \\ 3 & 1 \end{vmatrix} = (1)(1) - (1)(3) = -2 \qquad \text{(XIII.10)}$$

The i j "cofactor" $C_{i\,j}$ of **A** is $| \mathbf{A}_{i\,j} |$ multiplied by $(-1)^{i+j}$, or

$$C_{i\,j} = (-1)^{i+j} \, | \mathbf{A}_{i\,j} | \qquad \text{(XIII.11)}$$

or simply the signed minor determinant.

The recursive formula for | **A** | is obtained by "expanding" along a row or column of **A**, multiplying the relevant cofactors by the pivot elements of **A**. For example, for a 3 x 3 matrix **A** given by eq.(XIII.5) which is expanded along its first row,

$$| \mathbf{A} | = a_{11} C_{11} + a_{12} C_{12} + a_{13} C_{13} \qquad \text{(XIII.12a)}$$

$$= a_{11} | \mathbf{A}_{11} | - a_{12} | \mathbf{A}_{12} | + a_{13} | \mathbf{A}_{13} | \qquad \text{(XIII.12b)}$$

The method is recursive, because it expresses an n-th order determinant as the signed sum of (n-1)-th order determinants.

For a larger example, consider

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 0 & 1 \\ 3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad \text{(XIII.13)}$$

Using R,

```
> A<- matrix(c(2,1,3,0, 1,2,0,1, 3,1,0,0, 0,0,1,1), ncol=4, byrow=TRUE)
> A
     [,1] [,2] [,3] [,4]
[1,]    2    1    3    0
[2,]    1    2    0    1
[3,]    3    1    0    0
[4,]    0    0    1    1

> det(A)
[1] -16
```

The same result is obtain in Microsoft Excel using the cell formula

=MDETERM({2,1,3,0;1,2,0,1;3,1,0,0;0,0,1,1})

Note that if **A** is diagonal or triangular, | **A** | is simply the product of its diagonal elements.

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

XIV. APPLICATIONS OF INVERSES & DETERMINANTS IN STATISTICS:

Consider a sample of data of data measured on 2 variables for 4 subjects, and represented by the matrix

$$\mathbf{X} = \begin{bmatrix} 4 & 1 \\ 3 & 2 \\ 2 & 3 \\ 1 & 6 \end{bmatrix} \qquad N = 4,\ p = 2 \qquad \text{(XIV.1)}$$

which corresponds to the (row vector) measurements on the 4 experimental units of

$$\mathbf{x}_1 = [4 \quad 1] \qquad \text{(XIV.2a)}$$

$$\mathbf{x}_2 = [3 \quad 2] \qquad \text{(XIV.2b)}$$

$$\mathbf{x}_3 = [2 \quad 3] \qquad \text{(XIV.2c)}$$

$$\mathbf{x}_4 = [1 \quad 6] \qquad \text{(XIV.2d)}$$

The sample average is then

$$\mathbf{m} = \Sigma\, \mathbf{x}_i \ / \ N \qquad i = 1, 2, 3, 4 \qquad \text{(XIV.3a)}$$

$$= \{ [\ 4 \ \ 1\ ] + [\ 3 \ \ 2\ ] + [\ 2 \ \ 3\ ] + [\ 1 \ \ 6\ ] \} / 4 \qquad \text{(XIV.3b)}$$

$$= [\ 10 \ \ 12\ ] / 4 \qquad \text{(XIV.3c)}$$

$$= [\ 2.5 \ \ 3.0] \qquad \text{(XIV.3d)}$$

The sample covariance matrix is given by

$$\mathbf{S} = \Sigma\, (\mathbf{x}_i - \mathbf{m})' \, (\mathbf{x}_i - \mathbf{m}) \ / \ (N - 1) \qquad i = 1, 2, 3, 4 \qquad \text{(XIV.4)}$$

or

$$S_{j\,k} = \Sigma\, (x_{i\,j} - m_j)\, (x_{i\,k} - m_k) \ / \ (N - 1) \qquad j = 1,2 \ \ k = 1,2 \qquad \text{(XIV.5)}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

For the given data matrix **X**, this corresponds to

$$\mathbf{S} = \{ ([\,4\ \ 1\,] - [\,2.5\ \ 3.0\,])' ([\,4\ \ 1\,] - [\,2.5\ 3.0\,]) + ([\,3\ \ 2] - [\,2.5\ \ 3.0\,])' ([\,3\ \ 2] - [\,2.5\ \ 3.0\,]) + ([\,2\ \ 3] - [\,2.5\ \ 3.0\,])' ([\,2\ \ 3] - [\,2.5\ \ 3.0\,]) + ([\,1\ \ 6\,] - [\,2.5\ 3.0\,])' ([\,1\ \ 6\,] - [\,2.5\ \ 3.0\,]) \} / (4 - 1) \qquad \text{(XIV.6a)}$$

$$= \{ \begin{bmatrix} 1.5 \\ -2 \end{bmatrix} [\,1.5\ \ -2\,] + \begin{bmatrix} 0.5 \\ -1 \end{bmatrix} [\,0.5\ \ -1\,] + \begin{bmatrix} -0.5 \\ 0 \end{bmatrix} [\,-0.5\ \ 0\,] + \begin{bmatrix} -1.5 \\ 3 \end{bmatrix} [\,-1.5\ \ 3\,]\ \} / 3 \qquad \text{(XIV.6b)}$$

$$= \{ \begin{bmatrix} 2.25 & -3 \\ -3 & 4 \end{bmatrix} + \begin{bmatrix} 0.25 & -0.5 \\ -0.5 & 1 \end{bmatrix} + \begin{bmatrix} 0.25 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 2.25 & -4.5 \\ -4.5 & 9 \end{bmatrix} \} / 3 \qquad \text{(XIV.6c)}$$

$$= \begin{bmatrix} 5 & -8 \\ -8 & 14 \end{bmatrix} / 3 \qquad \text{(XIV.6d)}$$

$$= \begin{bmatrix} 5/3 & -8/3 \\ -8/3 & 14/3 \end{bmatrix} \qquad \text{(XIV.6e)}$$

Note how **S** is symmetric (i.e., equal to its transpose, see Section XVIII) and each of the elements going into it in eq.(XIV.6.c) are symmetric.

In summary, the sample mean vector **m** and sample covariance matrix **S** are

$$\mathbf{m} = [\,2.5\ \ 3.0] \qquad \text{(XIV.7a)}$$

$$\mathbf{S} = \begin{bmatrix} 5/3 & -8/3 \\ -8/3 & 14/3 \end{bmatrix} \qquad \text{(XIV.7b)}$$

(You should carry out the calculation of the elements of **m** and **S** in the way you are usually accustomed, and verify the same numerical values are obtained.)

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Reproducing these calculations using R:

```
> X<- matrix(c(4,1, 3,2, 2,3, 1,6), ncol=2, byrow=TRUE) #data
> X
     [,1] [,2]
[1,]    4    1
[2,]    3    2
[3,]    2    3
[4,]    1    6

> m<- colMeans(X) #sample mean vector
> m
[1] 2.5 3.0

> d<- t(t(X)-m)     #show deviation matrix
> d
     [,1] [,2]
[1,]  1.5   -2
[2,]  0.5   -1
[3,] -0.5    0
[4,] -1.5    3

> d[1,]%o%d[1,]  #show sample element
      [,1] [,2]
[1,]  2.25   -3
[2,] -3.00    4

> d[2,]%o%d[2,]
      [,1] [,2]
[1,]  0.25 -0.5
[2,] -0.50  1.0

> d[3,]%o%d[3,]
     [,1] [,2]
[1,] 0.25    0
[2,] 0.00    0

> d[4,]%o%d[4,]
      [,1] [,2]
[1,]  2.25 -4.5
[2,] -4.50  9.0

> S<- ( d[1,]%o%d[1,] + d[2,]%o%d[2,] + d[3,]%o%d[3,] + d[4,]%o%d[4,] ) / 3
#sample covariance matrix
> S
          [,1]      [,2]
[1,]  1.666667 -2.666667
[2,] -2.666667  4.666667

> cov(X)            #get same matrix from cov() function
          [,1]      [,2]
[1,]  1.666667 -2.666667
[2,] -2.666667  4.666667
```

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

To obtain the sample correlation matrix **R**, we divide each row and column element of **S** by the square root of the relevant diagonal element of **S** (i.e., the variance of the relevant variable), or

$$R_{jk} = S_{jk} \; / \; \sqrt{\{ S_{jj} \; S_{kk} \}} \qquad \text{(XIV.8)}$$

or

$$\mathbf{R} = (\mathbf{V}^{1/2})^{-1} \; \mathbf{S} \; (\mathbf{V}^{1/2})^{-1} \qquad \text{(XIV.9)}$$

where **V** is the matrix with the same principal diagonal as **S** with 0's elsewhere

$$\mathbf{V} = \mathrm{diag}(\mathbf{S}) \qquad \text{(XIV.10a)}$$

$$= \begin{bmatrix} 5/3 & 0 \\ 0 & 14/3 \end{bmatrix} \qquad \text{(XIV.10b)}$$

and

$$(\mathbf{V}^{1/2})^{-1} = \begin{bmatrix} 0.7746 & 0 \\ 0 & 0.4629 \end{bmatrix} \qquad \text{(XIV.11)}$$

Using eq.(XIV.7b) and eq.(XIV.11) in eq.(XIV.9), the sample correlation matrix **R** is given by

$$\mathbf{R} = \begin{bmatrix} 0.7746 & 0 \\ 0 & 0.4629 \end{bmatrix} \begin{bmatrix} 5/3 & -8/3 \\ -8/3 & 14/3 \end{bmatrix} \begin{bmatrix} 0.7746 & 0 \\ 0 & 0.4629 \end{bmatrix} \qquad \text{(XIV.12a)}$$

$$= \begin{bmatrix} 0.7746 & 0 \\ 0 & 0.4629 \end{bmatrix} \begin{bmatrix} 1.2910 & -1.2344 \\ -2.0656 & 2.1602 \end{bmatrix} \qquad \text{(XIV.12b)}$$

$$= \begin{bmatrix} 1.0000 & -0.9562 \\ -0.9562 & 1.0000 \end{bmatrix} \qquad \text{(XIV.12c)}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

Using R to calculate the sample correlation matrix:

```
> V<- diag(diag(S), ncol=2, nrow=2)   #make 2x2 matrix with same diagonal as S
> V
         [,1]     [,2]
[1,] 1.666667 0.000000
[2,] 0.000000 4.666667

> solve(sqrt(V))    #inv(sqrt(V))
          [,1]      [,2]
[1,] 0.7745967 0.0000000
[2,] 0.0000000 0.4629100

> W<- diag(1/sqrt(diag(S)), ncol=2, nrow=2)   #less computationally intensive
> W
          [,1]      [,2]
[1,] 0.7745967 0.0000000
[2,] 0.0000000 0.4629100

> R<- W %*% S %*% W     #compute correlation matrix
> R
           [,1]       [,2]
[1,]  1.0000000 -0.9561829
[2,] -0.9561829  1.0000000

> cor(X)            #directly from data
           [,1]       [,2]
[1,]  1.0000000 -0.9561829
[2,] -0.9561829  1.0000000
```

The normal distribution for a single variable has a probability density function of

$$f(x; \mu, \sigma^2) \quad = \quad \exp \{ -(x - \mu)2 / (2 \sigma^2) \} / \sqrt{(2 \pi \sigma^2)} \qquad \text{(XIV.13)}$$

where μ is the population mean and σ^2 the population variance.

In the multivariate case, there are p different variables combined into $\mathbf{x}$, and there are p different variances plus all of the covariances between them. These are represented by the covariance matrix $\mathbf{\Sigma}$ whose i,j-th element is

$$\Sigma_{i\,j} \quad = \quad E[(x_i - \mu_i) (x_j - \mu_j)] \qquad \text{(XIV.14)}$$

The multivariate normal distribution, used widely in multivariate statistics, has a probability density function of

$$f(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma}) \quad = \quad \exp\{ - (\mathbf{x} - \boldsymbol{\mu})^t \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})/2 \} / \{ (2 \pi)^{p/2} | \mathbf{\Sigma} |^{1/2} \} \qquad \text{(XIV.15)}$$

Note the occurrence of $\mathbf{\Sigma}^{-1}$ and $| \mathbf{\Sigma} |$ in eq.(XIV.15).

Consider the sample estimate $\mathbf{S}$ of the population variance $\mathbf{\Sigma}$, based on n data $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$, with elements $S_{i\,j}$ given by:

$$S_{i\,j} \quad = \quad \Sigma (x_{k\,i} - m_i) (x_{k\,j} - m_j) / (n - 1) \qquad \text{(XIV.16)}$$

Copyright 2009,2010,2011,2012,2013 by Robert A LaBudde, all rights reserved.

where

$$m_i = \Sigma x_{k\,i} / n \qquad \text{(XIV.17a)}$$

$$m_j = \Sigma x_{k\,j} / n \qquad \text{(XIV.17b)}$$

are the sample means.

Then | **S** | is called the "generalized sample variance". It conveniently supplies a single numerical value expressing the dispersion of the multivariate sample, and is proportional to the square of the volume enclosed by the n vector deviations from the vector mean in the data. If | **S** | = 0, these deviations are linearly dependent and degenerate.