Systems of Linear Equations

Statistics 427: R Programming

Module 12

2020

Matrix Representation

By precalculus, you may have had an introduction to systems of linear equations.

$$-x_1 + 4x_2 = 8$$
$$3x_1 + 6x_2 = 30$$

Here x_1 and x_2 are unknown quantities. You could use the notation x and y rather than x_1 and x_2 (six and one half dozen of the other). However, when systems of equations have more than 2 unknowns (more than 2 dimensional coordinates), you will need more letters in the alphabet and it is easier to use x with subscripts.

Definition

The set of points with coordinates (x_1, x_2) on a two-dimensional Cartesian graph that satisfy the first equation is a line. Similarly, the set of points with coordinates x_1, x_2 on a two-dimensional Cartesian graph that satisfy the second equation is also a line, different from the first.

Form of equations

We can arrange the two equations a bit to see better that the equations are in a familiar form $y = a + bx$. Other variations you may have seen are (the classic) $y = mx + b$, $y = \beta_0 + \beta_1 x_i$, the list goes on. The basic form for the equations are

$$x_2 = (intercept) + (slope)x_1$$

Rearranging equations

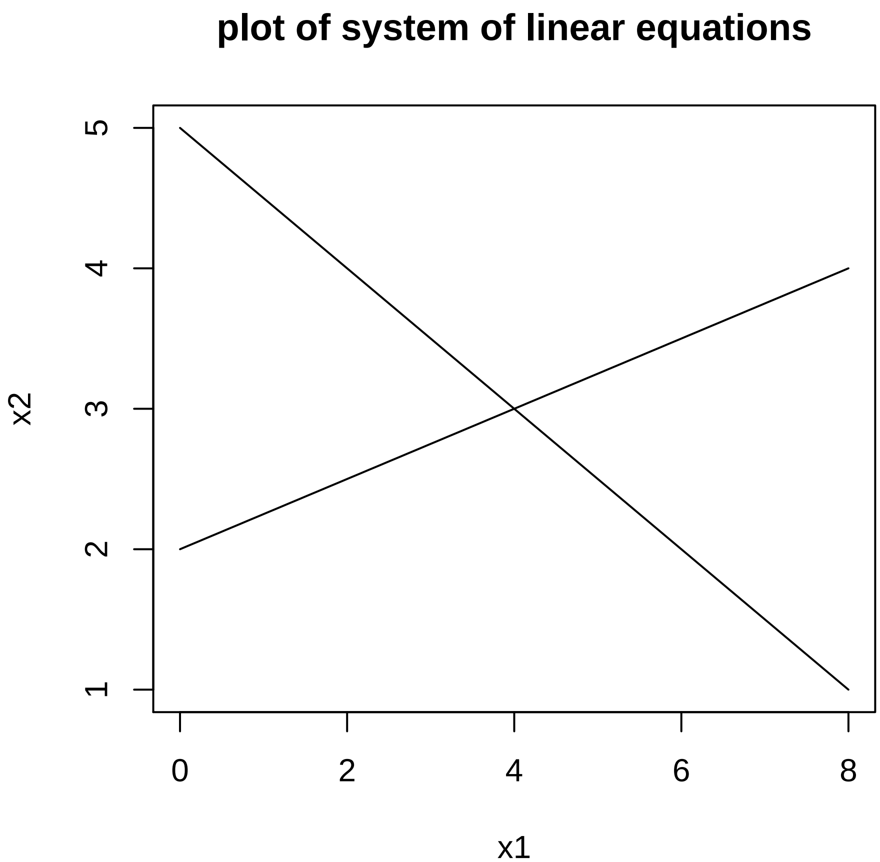
$$x_2 = 2 + \frac{1}{4}x_1$$

$$x_2 = 5 - \frac{1}{2}x_1$$

By a *solution to the system of linear equations*, it means a set of values of x_1, x_2 that satisfy both equations simultaneously (the set of points will lie on both lines at the same time). Such a solution can happen in two ways: (a) the lines intersect at a point or (b) the lines coincide. If the lines are parallel, they will never meet and no solution exists.

New plot function for system of linear equations

```r
x1=(0:100)*8/100 # values from 0 to 8
x2.a=2+(1/4)*x1 # first line x2 values
x2.b=5-(1/2)*x1 # second line x2 values
x2=cbind(x2.a,x2.b) # combine all x2 values as columns in matrix
matplot(x1,x2,type='l',lty=c(1,1),col=c(1,1)); title('plot of system of linear equations')
```



Discussion of graph

The first line had a vertical axis intercept of 2 and an increasing slope; the second has an intercept of 5 and a decreasing slope. The lines intersect at a point, $(4, 3)$. The dreaded part of algebra comes next: you can use that solution to double check it by plugging in the point and making sure the equations stay true.

These equations, however, can be expressed in terms of matrices. Use the coefficients of x_1 and x_2 and form the original two equations into a matrix.

Convert data into matrices

$$-x_1 + 4x_2 = 8$$

$$3x_1 + 6x_2 = 30$$

$$\begin{bmatrix} -1 & 4 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 30 \end{bmatrix}$$

Matrix equation

If the coefficient matrix is denoted by $\mathbf{A}$, the column vector of unknowns by $\mathbf{x}$, and the column of vector constants by $\mathbf{c}$, we can symbolically write the matrix equation as

$$\mathbf{Ax} = \mathbf{c}$$

Using "regular" algebra to solve simultaneous linear equations can get tedious. Solving for one variable algebraically or by elimination, then substituting the values back into the system and solving for another unknown variable, later, rinse, repeat ad nauseum. No thank you (and I ♡ algebra!)

There is no spoon

(sorry. . . just couldn't help myself)

Can we just divide both sides of the matrix equation by $\mathbf{A}$? No, because there is no such thing as matrix division. But we can multiply both sides by the inverse of the matrix $\mathbf{A}$. Think of it like multiplying a value x by its reciprocal.

$$x\left(\frac{1}{x}\right) = 1$$

Matrix Inverse

Something akin to matrix division for square matrices (those that have the same number of rows and columns) can be defined by analogy to ordinary division of real numbers. Ordinary division of real numbers is actually multiplication. For multiplication of real numbers, the *multiplicative identity* is the number 1, that is any real number a multiplied by 1 is just a: $a(1) = a$. The *reciprocal* or *inverse* of the number a is another real number (call it b) such that if you multiply it by a you get the multiplicative identity: $ba = ab = 1$. We know this number b as $\frac{1}{a}$ or a^{-1}.

Reciprocals do not exist for all real numbers (such as 0).The key idea for extending division to matrices is that division by a real number a is multiplication by a^{-1}

Identity matrix

For matrix multiplication, a special square matrix called the *identity matrix* is the multiplicative identity. An identity matrix with k rows and k columns is universally denoted with the letter $\mathbf{I}$ and takes on the form:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Identity matrix

It is a square matrix with 1s on the diagonal from upper left to lower right (also frequently called the main diagonal) and 0s elsewhere. If $\mathbf{C}$ is any matrix (square or otherwise), and an identity matrix of the right size is constructed so that the matrix multiplication is defined, then $\mathbf{CI} = \mathbf{C}$ (and $\mathbf{IC} = \mathbf{C}$), where the columns of $\mathbf{I}$ is equal to the number of rows of $\mathbf{C}$.

The inverse of a square matrix $\mathbf{A}$, denoted as $\mathbf{A^{-1}}$, is a square matrix of the same size as $\mathbf{A}$ that produces an identity matrix when pre- or post-multiplied by $\mathbf{A}$. That is,

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

Tiny, little baby example

Let

$$\mathbf{A} = \begin{bmatrix} -1 & 4 \\ 3 & 6 \end{bmatrix}$$

Tiny, little baby example

Find $\mathbf{A^{-1}}$, the inverse of $\mathbf{A}$. Here is how to do that with a nice 2X2 example. After this, we use `R`.

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

$$\mathbf{A^{-1}} = \frac{\mathbf{1}}{\mathbf{-6 - 12}} \begin{bmatrix} -1 & 4 \\ 3 & 6 \end{bmatrix} = -\frac{\mathbf{1}}{\mathbf{18}} \begin{bmatrix} 6 & -4 \\ -3 & -1 \end{bmatrix}$$

Tiny, little baby example

$$\mathbf{A^{-1}} = \begin{bmatrix} -\frac{6}{18} & \frac{4}{18} \\ \frac{3}{18} & \frac{1}{18} \end{bmatrix}$$

Matrix of the system

If we know or can calculate the inverse of $\mathbf{A}$, that will lead to the solution of the system of linear equations. Take the matrix of the system, $\mathbf{Ax} = \mathbf{c}$ and premultiply both sides of the equation by the inverse of $\mathbf{A}$.

$$\mathbf{A^{-1}Ax} = \mathbf{A^{-1}c}$$

Because of the associative property of matrix multiplication, on the left side we can do the $\mathbf{A^{-1}A}$ first

$$\mathbf{Ix} = \mathbf{A^{-1}c}$$

Matrix of the system

With the multiplicative identity, this is really

$$\mathbf{x} = \mathbf{A^{-1}c}$$

Or

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -\frac{6}{18} & \frac{4}{18} \\ \frac{3}{18} & \frac{1}{18} \end{bmatrix} \begin{bmatrix} 8 \\ 30 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

And this is the solution we came up with from the graph

Systems of linear equations

As mentioned before, systems of linear equations can have more than two unknown parameters (x_i) to solve for. A linear equation with three unknowns is the equation for a plane in three dimensions (3-D space). Three equations in three unknowns can have a unique point solution (think corner of a box or where two walls and a ceiling meet). Four equations in four unknowns, or k equations in k unknowns, can have unique point solutions as well. Linear equations with four or more unknowns are called hyperplanes and cannot be envisioned well in our three-dimensional experience. We can still describe things with a vector of four or more numbers is rather routine in science.

Systems of linear equations

However many unknowns there are in the system of equations, the system has a matrix representation of the form $\mathbf{Ax} = \mathbf{c}$. Concentration will be on systems in which the number of unknowns is the same as the number of equations, thus the matrices will be square matrices.

Proof. . . not

The result of the proof of the form $\mathbf{Ax} = \mathbf{c}$ being the form that holds for k unknowns with k equations is what we will look at. The proof for this is beyond the scope of this course; if interested, consider taking Math 330 Linear Algebra.

Result: The system of linear equations defined by $\mathbf{Ax} = \mathbf{c}$ where $\mathbf{A}$ is a $k \times k$ matrix and at least one of the elements of $\mathbf{c}$ is nonzero, has a unique point solution if an inverse matrix for $\mathbf{A}$ exists. The solution is then given by $\mathbf{x} = \mathbf{A^{-1}c}$

Basically, if an inverse exists, you can find a unique point solution.

Finally, how to do this in `R`

. . . almost. One new function to use here is `solve()` and it can be used two ways: (1) to solve a system of linear equations, provided a solution exists, or (2) to calculate the inverse of a square matrix, provided the inverse exists. Both tasks are accomplished within the constraints of floating point arithmetic (which creates some round-off error), and either might fail in situations (such as large sparse matrices) in which round-off errors tend to build up quickly.

Finally, how to do this in `R`!!!

```r
A=rbind(c(-1,4),c(3,6)) # by rows
c=c(8,30)
x=solve(A,c)
Ainv=solve(A)
x
```

```
[1] 4 3
```

Finally, how to do this in `R`!!!

```r
Ainv
```

```
           [,1]       [,2]
[1,] -0.3333333 0.22222222
[2,]  0.1666667 0.05555556
```

```r
A%*%Ainv # check to see of we get indentity matrix (or closeish)
```

```
     [,1]          [,2]
[1,]    1 -2.775558e-17
[2,]    0  1.000000e+00
```

```r
Ainv%*%c # same as x=solve(A,c)
```

```
     [,1]
[1,]    4
[2,]    3
```

Finally, how to do this in R!!!

```r
# function for fum
# s.se is for solve system of equations
s.se=function(A,c){
  x=solve(A,c); Ainv=solve(A)
  cat('',"Solution = ",x,'')
  results=list(x=x,Ainv=Ainv) # you can use $ to extract Ainv
}
B=rbind(c(3,4,-6),c(2,-5,17),c(-1,8,-4))
d=c(10,30,5)
s.se(B,d)
```

```
 Solution =  4.288889 2.1 1.877778
```

Finally, how to do this in R!!!

```r
s.se(B,d)$Ainv # s.se(B,d)[2] works as well (Ainv is 2nd element)
```

```
 Solution =  4.288889 2.1 1.877778

            [,1]       [,2]        [,3]
[1,]  0.25777778 0.07111111 -0.08444444
[2,]  0.02000000 0.04000000  0.14000000
[3,] -0.02444444 0.06222222  0.05111111
```

Inverse matrix that does not exist?

Not all square matrices have inverses. Unlike real numbers, there are many matrices without inverses. An inverse does not exist for the following matrix:

$$\mathbf{A} = \begin{bmatrix} -1 & 4 \\ -2 & 8 \end{bmatrix}$$

The two equations have the same slope but different vertical axis intercepts, meaning they are parallel and will never intersect at a unique point solution. A matrix with no inverse is said to be *singular*, and (attempted) solving a system defined by such a matrix (like dividing by zero); a matrix with an inverse is *nonsingular*

Fitting data

To find the equation of a line that basically passes somehow through the middle of the data might offer a decent prediction of the vertical axis variable based on the variable of the horizontal axis. By investigating only individual pairs of data points, you cannot get accurate enough ideas for what the vertical axis variable is doing. Looking at a sample of data is the way to attempt to estimate the line that "best fits" the data.

Applications of systems of linear equations

One of the most important and widely used applications of solving a system of linear equations is to find a good prediction line in just such situations where there is a strong linear relationship between the two quantities.

The horizontal axis variable is called the explanatory variable and the vertical axis variable is the response variable (the one to predict based on the explanatory variable; modeling y based on x).

The goal: use the line that predicts the data at hand the best.

Best fit line

Small prediction errors: using the concept *squared prediction error* for each observation. For the ith observation, take the value of y_i of the response variable, subtract the value $\hat{y}_i = b_1 + b_2 x_i$ predicted for that observation by a particular line calculated with the predictor variable and square the result: $\hat{y}_i^2 = (y_i - (b_1 + b_2 x_i))^2$. This squared prediction error is a measure of a lack of fit for that observation. Such a measure should magnify a large departure of the observed value of y from the predicted value by the line.

We can use the *sum of the squared errors* (SSE) given by:

$$SSE = \sum (y_i - \hat{y}_i)^2 = (y_1 - b_1 - b_2 x_1)^2 + (y_2 - b_1 - b_2 x_2)^2 + \cdots + (y_n - b_1 - b_2 x_n)^2$$

Best fit line

The criterion we can use is to pick the values of b_1 and b_2 that make the SSE as small as possible. The values of b_1 and b_2 that minimize the sum of squared errors are called the *least squares estimates* of b_1 and b_2. The least squares criterion for picking prediction equations is used extensively in science and business.

There is actually a unique pair of parameter estimate values (b_1 and b_2) that minimize the sum of the squared errors for *any* given dataset. Stated alternatively; one unique line minimizes SSE. With matrix calculations, obtaining the least squares estimate is relatively straightforward.

Response variable vector (matrix)

The observed values of y will make up one column vector.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Matrix of X

The next matrix, called $\mathbf{X}$, will have n rows and two columns, in which the elements of the first column are all 1s and the elements of the second column are the observations on the predictor variable x.

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

Transpose of X

The transpose of $\mathbf{X}$ will be the matrix formed when switching the rows and columns of the matrix (rows become columns and columns become rows). The notation for the transpose of $\mathbf{X}$ is $\mathbf{X}'$.

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

Unknowns

The column vector of the unknowns of the intercept and slope constants are denoted by $\mathbf{b}$

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Form of equation

The least squares estimates of the intercept and slope constants are a point solution to the system of linear equations given by

$$\mathbf{Ab} = \mathbf{c}$$

Where $\mathbf{A}$ is a square matrix formed by $\mathbf{A} = \mathbf{X}'\mathbf{X}$ and $\mathbf{c}$ is a column vector formed by $\mathbf{c} = \mathbf{X}'\mathbf{y}$. All we have to do is create the matrices and solve the system. The solution, denoted by $\hat{\mathbf{b}}$ if it exists, is given by $\mathbf{A}^{-1}\mathbf{c}$, or

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

.

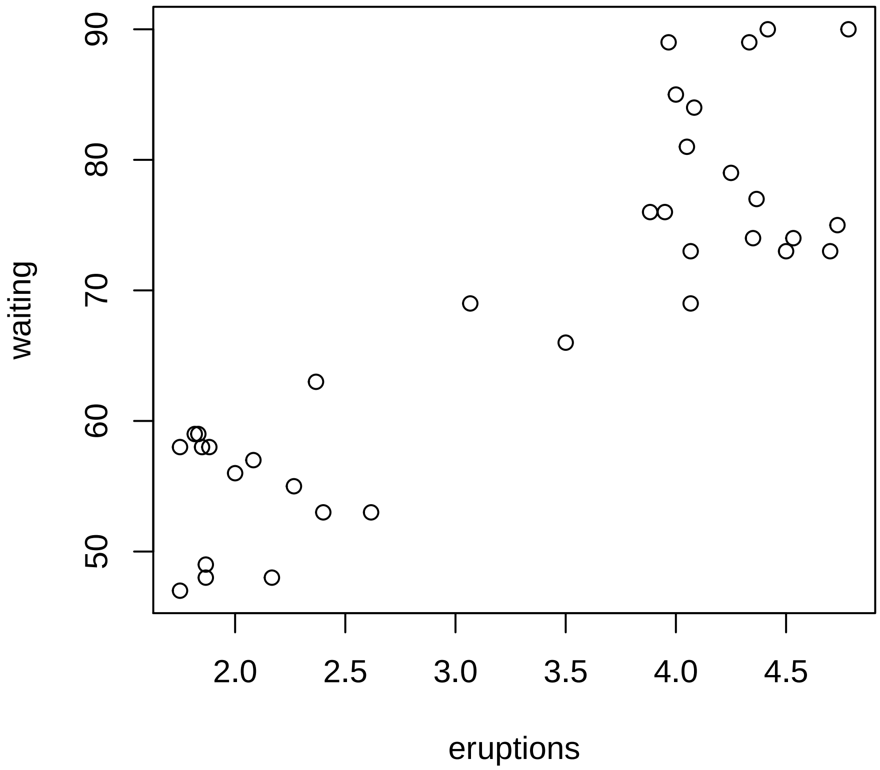
Old Faithful

```r
set.seed(1) # so you get the same sample I do
OF=data.frame(faithful)
faith=OF[sample(nrow(OF),35),]
head(faith)
```

```
    eruptions waiting
167     2.367      63
129     2.267      55
270     4.417      90
187     4.083      84
85      4.067      73
263     1.850      58
```

Old Faithful

```
with(faith,plot(waiting~eruptions))
```




```
x=faith$eruptions; y=faith$waiting
```

Least squares prediction

```
n=length(y) # sample size
X=matrix(1,n,2) # col1 has 1s
X[,2]=x # col2 has predictor variable x
xtx=t(X)%*%X # (X'X)
xty=t(X)%*%y # (X'y)
b=solve(xtx,xty)
b
```

```
         [,1]
[1,] 34.52443
[2,] 10.29334
```

Least squares prediction

```
plot(x,y,main='Old Faithful',xlab='Eruptions',ylab='Waiting')
ypred1=b[1]+b[2]*min(x)
ypred2=b[1]+b[2]*max(x) # use x min and max to find 2 points for line
ypred=rbind(ypred1,ypred2)
```

```
xval=rbind(min(x),max(x))
points(xval,ypred,type='l')
```