450A Political Methodology I
TA Section 7

Vincent Bauer and William Marble

November 11, 2016

Road Map

1. Midterm Review
    1.1 Small groups present Qs 1-4
2. Homework Review
3. Matrix Regression
    3.1 Motivation
    3.2 Matrix Properties
    3.3 Matrix Derivations

Homework Review

1. Just saying that there might be omitted variables is meaningless, give a substantive example. Imagine you are at a workshop and giving a comment.
2. When are we concerned about omitted variables affectng our causal interpretation? Is party/religiosity an omitted variable when estimating the causal effect of ngirls on AAUW score after including totchi?
3. Is R dropping experience from the regression wages $\sim$ age $+$ education $+$ experience because it is highly correlated (.95) with the other IVs?

Motivating the Matrix Approach

Why are we introducing this new language?
Compare bivariate regression with a simple multivariate regression.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$$

$$\hat{\beta}_1 = \frac{\text{Cov}(X_1, Y)}{\text{Var}(X_1)}$$

This is pretty managable, but then add even just one more variable

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

$$\hat{\beta}_1 = \frac{\text{Cov}(X_1, Y)\text{Var}(X_2) - \text{Cov}(X_2, Y)\,\text{Cov}(X_1, X_2)}{\text{Var}(X_1)\text{Var}(X_2) - \text{Cov}(X_1, X_2)^2}$$

Calculating all of these quantities is tedious and this is still a very simple case. You could easily include 10 additional variables in your research.

Motivating the Matrix Approach

Using matrices, we can calculate all of our coefficients at once using

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

Calculating our parameters for statistical inference is just as easy.

- We need the variance-covariance matrix for the coefficients

$$Var[\hat{\boldsymbol{\beta}}] = (\sigma^2\mathbf{I})(\mathbf{X}^T\mathbf{X})^{-1}$$

`vcov(lm)`

- Calculate σ^2 just as before.

$$\hat{\sigma}^2 = \frac{SSR}{n-(k+1} = \frac{\hat{u}^T\hat{u}}{n-(k+1)}$$

Matrix Properties

Let's review some basic matrix concepts

1. What is the difference between a scalar, a vector, and a matrix?
2. What is the identity matrix?
3. What is a square matrix?
4. What are the rules for the addition and subtraction of matrices, and when can we carry out this operation?
5. What is the operation for transposition?
6. What is the operation for scalar mulplication?
7. What are the dimensions of $\mathbf{X}$? What about $\mathbf{X^T X}$?

Matrix Multiplication

Matrix multiplication is more involved: Given an $(m \times n)$ matrix $\mathbf{A}$ and an $(n \times r)$ matrix $\mathbf{B}$, we can define the product $\mathbf{AB} = \mathbf{C}$ as the $(m \times r)$ matrix whose entries are defined by:

$$\sum_{k=1}^{n} a_{i,k}\, b_{k,j} \;=\; c_{i,j}$$

where $i = 1, ..., m$ and $j = 1, ..., r$.

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

$$\mathbf{C} \;=$$

Matrix Multiplication

Matrix multiplication is more involved: Given an $(m \times n)$ matrix $\mathbf{A}$ and an $(n \times r)$ matrix $\mathbf{B}$, we can define the product $\mathbf{AB} = \mathbf{C}$ as the $(m \times r)$ matrix whose entries are defined by:

$$\sum_{k=1}^{n} a_{i,k}\, b_{k,j} \;=\; c_{i,j}$$

where $i = 1, ..., m$ and $j = 1, ..., r$.

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

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} -2 \cdot 3 + 1 \cdot 2 + 3 \cdot 1 & -2 \cdot -2 + 1 \cdot 4 + 3 \cdot -3 \\ 4 \cdot 3 + 1 \cdot 2 + 6 \cdot 1 & 4 \cdot -2 + 1 \cdot 4 + 6 \cdot -3 \end{bmatrix} \\ &= \begin{bmatrix} -1 & -1 \\ 20 & -22 \end{bmatrix} \end{aligned}$$

$\mathbf{A}$ is (2×3) and $\mathbf{B}$ is (3×2), so $\mathbf{C}$ is (2×2).

Matrix Multiplication Properties

- Matrix multiplication is associative and distributive

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC}) \text{ and } \boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC}$$

- But it is not commutative, order **matters**

$$\boldsymbol{AB} \neq \boldsymbol{BA}$$

- Transposition also distributes

$$(\boldsymbol{AB})^{\boldsymbol{T}} = \boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{A}^{\boldsymbol{T}} \text{ and } (\boldsymbol{A}+\boldsymbol{B})^{\boldsymbol{T}} = \boldsymbol{A}^{\boldsymbol{T}}+\boldsymbol{B}^{\boldsymbol{T}}$$

- Multiplication with the inverse equals the identity matrix

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

- The identity matrix is commutative, associative, and distributive

$$\boldsymbol{AIB} = \boldsymbol{IAB} = \boldsymbol{ABI} = \boldsymbol{AB}$$

Matrix Multiplication in R

```
a <- matrix(c(1,4,3,2), ncol=2); a

##      [,1] [,2]
## [1,]    1    3
## [2,]    4    2

b <- matrix(c(1,8, 9, 4), ncol=2); b

##      [,1] [,2]
## [1,]    1    9
## [2,]    8    4

a + b

##      [,1] [,2]
## [1,]    2   12
## [2,]   12    6
```

Matrix Multiplication in R

```
a

##      [,1] [,2]
## [1,]    1    3
## [2,]    4    2

b

##      [,1] [,2]
## [1,]    1    9
## [2,]    8    4
```

```
a[1,] * b[,1]

## [1]  1 24

sum(a[1,] * b[,1] )

## [1] 25

a %*% b

##      [,1] [,2]
## [1,]   25   21
## [2,]   20   44
```

Matrix Multiplication in R

```
a %*% b

##      [,1] [,2]
## [1,]   25   21
## [2,]   20   44
```

```
sum(a[1,] * b[,1] )

## [1] 25

sum(a[1,] * b[,2] )

## [1] 21

sum(a[2,] * b[,1] )

## [1] 20

sum(a[2,] * b[,2] )

## [1] 44
```

Matrix Inversion

In practice you will ask R to invert matrices for you.

```
a

##      [,1] [,2]
## [1,]    1    3
## [2,]    4    2

solve(a)

##      [,1] [,2]
## [1,] -0.2  0.3
## [2,]  0.4 -0.1
```

But what is R doing?

We can compute the inverse of a 2x2 matrix by hand.

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ then } \mathbf{A}^{-1} = \frac{1}{a \cdot d - b \cdot c} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}, \text{ then } \mathbf{A}^{-1} = \frac{1}{1 \cdot 2 - 3 \cdot 4} \begin{bmatrix} 2 & -3 \\ -4 & 1 \end{bmatrix}$$

$$\frac{-1}{10} \begin{bmatrix} 2 & -3 \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} -.2 & .3 \\ .4 & -.1 \end{bmatrix}$$

Note: the quantity $\frac{1}{a \cdot d - b \cdot c}$ is called the determinant.

But once matrices get bigger than this the computation is much harder. R is probably using a fancy algorithm instead of the rule we used.

Matrix Application Toy Example

Lets say we had receipts from a job-talk dinner where grad students and faculty went to a restaurant with drinks and food. There are two drink options, whiskey and wine, and two food options, burgers and fish.

We want to know how many students and how many faculty went to the dinner. Our receipts only tell us the total amount spent on drinks (220) and on food (260). We check their menu online and learn that the prices are, respectively, 8,6,8,10. We know from qualitative case studies that all students order whiskey and burgers and that all faculty order wine and fish.

This sets up the following equations

$$x * 8 + y * 6 = 220$$

$$x * 8 + y * 10 = 260$$

There are two (independent) equations and two unknowns, so we can solve this system.

We could solve this system by plugging one equation into the other

First, rearrange the second equation

$$260 - y * 10 = x * 8$$

Plug this into the first

$$260 - y * 10 + y * 6 = 220$$

Solve for y

$$y = 10$$

Plug back in

$$x * 8 + 10 * 6 = 220$$
$$x * 8 + 60 = 220$$
$$x * 8 = 160$$
$$x = 20$$

But we can also solve it with matrices. Here's our rewritten system.

$$\begin{bmatrix} x & y \end{bmatrix} \cdot \begin{bmatrix} 5 & 8 \\ 6 & 10 \end{bmatrix} = \begin{bmatrix} 220 \\ 260 \end{bmatrix}$$

Now isolate the quantities of interest

$$\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} 5 & 8 \\ 6 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 220 \\ 260 \end{bmatrix}$$

Take the inverse (determinant = .5)

$$\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} 5 & -4 \\ -3 & 2.5 \end{bmatrix} \begin{bmatrix} 220 \\ 260 \end{bmatrix}$$

Matrix multiply

$$\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} 5 \cdot 220 - 4 \cdot 260 \\ -3 \cdot 220 + 2.5 \cdot 260 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

Derivation of OLS

$$S(\beta, \boldsymbol{X}, \boldsymbol{y}) = \hat{\boldsymbol{u}}^{\boldsymbol{T}} \hat{\boldsymbol{u}} = SRR, \text{ we want to minimize this}$$
$$= (\boldsymbol{y} - \boldsymbol{X}\beta)^{\boldsymbol{T}} (\boldsymbol{y} - \boldsymbol{X}\beta)$$

Apply the transpose rule twice

$$= (\boldsymbol{y}^{\boldsymbol{T}} - \beta^{\boldsymbol{T}} \boldsymbol{X}^{\boldsymbol{T}})(\boldsymbol{y} - \boldsymbol{X}\beta)$$

Distribute, making sure to keep the order the same

$$= \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - \beta^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} - \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X}\beta + \beta^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\beta$$

Notice $\beta^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y}$ is a (1 x k)(k x n)(n x 1) = (1x1) matrix so $\boldsymbol{A} = \boldsymbol{A}^{\boldsymbol{T}}$

$$= \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - \boldsymbol{y}^{\boldsymbol{T}}(\boldsymbol{X}\beta) - \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X}\beta + \beta^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\beta$$
$$= \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y} - 2\boldsymbol{y}^{\boldsymbol{T}}(\boldsymbol{X}\beta) + \beta^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\beta$$

Some quick definitions before we continue

1. Derivatives work almost exactly the same for matrices.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + c, \text{ then } \frac{d\boldsymbol{y}}{d\boldsymbol{X}} = \boldsymbol{\beta}^T$$

   Notice the transpose.

2. We call the column vector (k x 1) of partial derivatives the gradient

$$\frac{d\boldsymbol{y}}{d\boldsymbol{X}} = \begin{bmatrix} dy/dX_1 = \beta_1 \\ dy/dX_2 = \beta_2 \\ \cdots \\ dy/dX_k = \beta_k \end{bmatrix}$$

   Note that these are columns of $\boldsymbol{X}$, not rows of $\boldsymbol{X}$.

3. We call the matrix (k x k) of 2nd-order partial derivatives the Hessian
4. Squares are also basically the same

$$\frac{dy}{dA}\ \boldsymbol{A^T A} = \boldsymbol{2A}, \text{ and }, \frac{dy}{dA}\ \boldsymbol{A^T X A} = \boldsymbol{2XA}$$

$$S(\beta, \boldsymbol{X}, \boldsymbol{y}) = \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{y}^T(\boldsymbol{X}\beta) + \beta^T\boldsymbol{X}^T\boldsymbol{X}\beta$$

$$\frac{dS(\beta, \boldsymbol{X}, \boldsymbol{y})}{d\beta} = -2\boldsymbol{X}^T\boldsymbol{y} + 2\boldsymbol{X}^T\boldsymbol{X}\beta$$

Set equal to zero

$$0 = -2\boldsymbol{X}^T\boldsymbol{y} + 2\boldsymbol{X}^T\boldsymbol{X}\tilde{\beta}$$

$$2\boldsymbol{X}^T\boldsymbol{X}\tilde{\beta} = 2\boldsymbol{X}^T\boldsymbol{y}$$

Multiply by the inverse, making sure to keep the order consistent

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{X})\tilde{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{y})$$

$$\boldsymbol{I}\tilde{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{y})$$

Unbiasedness of $\hat{\beta}$

$$
\begin{aligned}
\hat{\beta} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\beta + \mathbf{u}) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}) \\
&= I\beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u} \\
E[\hat{\beta}|\mathbf{X}] &= \beta + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E[\mathbf{u}|\mathbf{X}] \\
&= \beta \\
E[E[\hat{\beta}|\mathbf{X}]] &= E[\beta] = \beta
\end{aligned}
$$

Variance of $\hat{\beta}$

$$\begin{aligned}
V[\hat{\beta}|\mathbf{X}] &= V[\beta|\mathbf{X}] + V[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}|\mathbf{X}] \\
&= 0 + V[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}|\mathbf{X}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T V[\mathbf{u}|\mathbf{X}]((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T V[\mathbf{u}|\mathbf{X}]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2\mathbf{I}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}
\end{aligned}$$

This yields a (k x k) variance-covariance matrix for $\hat{\beta}$ (including intercept in k).

Calculate σ^2 with $\dfrac{SSR}{n-k} = \dfrac{\hat{u}^T\hat{u}}{n-k}$.