

POLS/CS&SS 503:
Advanced Quantitative Political Methodology

MATRIX ALGEBRA, LINEAR REGRESSION

April 7, 2015

Jeffrey B. Arnold



Agenda

- Linear regression as finding a “best” line
- Linear regression as the conditional expectation function
- How linear regression relates to the normal distribution

What is regression?

Regression

distribution of a **response** (outcome) variable Y — or summary of that distribution — as a function of **explanatory** variables X_1, \dots, X_k .

Ordinary Least Squares

Finds a $\hat{Y} = \mathbf{X}\mathbf{B}$ that minimizes $\sum (Y_i - \hat{Y}_i)^2$. This estimates a linear conditional expectation function $E(Y|X_1, \dots, X_k)$.

OLS Objective Function

One X

Find the line

$$\hat{Y} = A + BX$$

such that

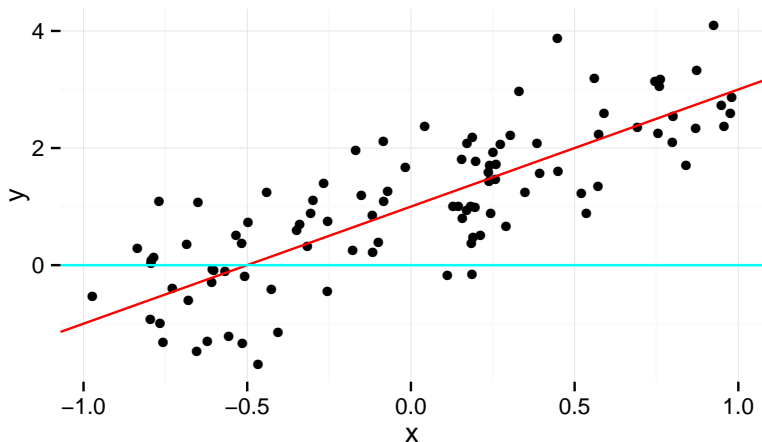
$$A, B = \arg \min_{A, B} S(A, B)$$

where

$$S(A, B) = \sum_i E_i^2 = \sum_i (Y_i - \hat{Y}_i)^2 = \sum_i (Y_i - A - BX_i)^2$$

How do we minimize this?

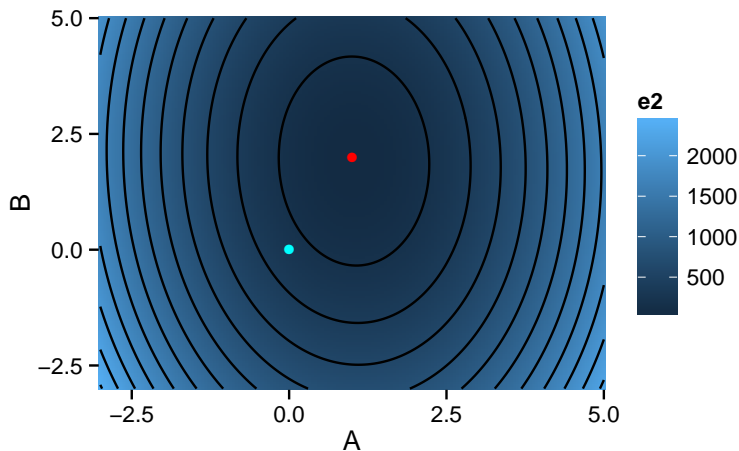
What does the OLS objective function look like?



Data generated by $Y_i = 1 + 2X_i + E_i$. Lines are $A = 1, B = 2$, and $A = 0, B = 0$.

$\sum E_i^2$ as a function of A and B

Least squares is the minimum of this function



Finding the best A, B in Least Squares

One X

To minimize, set partial derivatives equal to 0 and solve:

$$\frac{\partial S(A, B)}{\partial A} = \sum (-1)(2)(Y_i - A - BX_i) = 0$$

$$\frac{\partial S(A, B)}{\partial B} = \sum (-X_i)(2)(Y_i - A - BX_i) = 0$$

Rearrange to get

$$A = \bar{Y} - B\bar{X}$$

$$B = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{c(X, Y)}{v(X)}$$

Implications of the OLS Solution

Least squares A and B

$$A = \bar{Y} - B\bar{X}$$

$$B = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{c(X, Y)}{v(X)}$$

- \bar{X}, \bar{Y} is in the regression line
- $\sum X_i E_i = 0$

$$\begin{aligned}\sum X_i E_i &= \sum X_i (Y_i - A - B X_i) \\ &= \sum X_i Y_i - A \sum X_i - B \sum X_i^2 = 0\end{aligned}$$

- $\sum \hat{Y}_i E_i = 0$
- Errors E uncorrelated with \hat{Y} and X

OLS Objective Function

Multiple X

Find plane

$$Y = A + B_1X_1 + B_2X_2 + \cdots + B_kX_k$$

such that

$$A, B_1, \dots, B_k = \arg \min_{A, B_1, \dots, B_k} S(A, B_1, \dots, B_k)$$

where

$$\begin{aligned} S(A, B_1, \dots, B_k) &= \sum_i E_i^2 = \sum_i (Y_i - \hat{Y}_i)^2 \\ &= \sum_i (Y_i - A - \sum_{j=1}^k B_j X_{i,j})^2 \end{aligned}$$

How do we minimize this?

Finding the best A, B in Least Squares Regression

Multiple X

Set partial derivatives equal to 0 and solve system of equations for

$$\frac{\partial S(A, B_1, B_2, \dots, B_k)}{\partial A} = \sum (-1)(2)(Y_i - A - BX_i) = 0$$

$$\frac{\partial S(A, B_1, B_2, \dots, B_k)}{\partial B_1} = \sum (-X_{i,1})(2)(Y_i - A - B_1X_{i,1} - \dots - B_kX_{i,k}) = 0$$

$$\vdots = \vdots$$

$$\frac{\partial S(A, B_1, B_2, \dots, B_k)}{\partial B_k} = \sum (-X_{i,k})(2)(Y_i - A - B_1X_{i,1} - \dots - B_kX_{i,k}) = 0$$

Not as easy ...

Linear Regression in Matrix Form

Scalar representation

$$Y_i = B_0 + B_1 X_{i,1} + B_2 X_{i,2} + \dots B_k X_{i,k} + E_i$$

Equivalent matrix representation

$$\underset{n \times 1}{\mathbf{y}} = \underset{n \times (k+1)}{\mathbf{X}} \underset{(k+1) \times 1}{\mathbf{b}} + \underset{n \times 1}{\mathbf{e}}$$

or

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \cdots & X_{k,1} \\ 1 & X_{1,2} & X_{2,2} & \cdots & X_{k,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1,n} & X_{2,n} & \cdots & X_{k,n} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{bmatrix} + \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix}$$

Linear Regression in Matrix Form

Objective Function

The linear regression is

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

Want to find the \mathbf{b} that minimizes the squared errors:

$$\arg \min_{\mathbf{b}} S(\mathbf{b})$$

where

$$\begin{aligned} S(\mathbf{b}) &= \sum E_i^2 = \mathbf{e}'\mathbf{e} \\ &= (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \end{aligned}$$

Why does \mathbf{e} need to be transposed?

Linear Regression in Matrix Form

Transpose of Sums

$$(A + B)' = A' + B'$$

$$\left(\begin{bmatrix} 10 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right)' = ?$$
$$? = ?$$

Linear Regression in Matrix Form

Transpose of a product

$$(XB)' = B'X'$$

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

$$? = ?$$

Simplify $e'c$

$$\begin{aligned}e'e &= (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \\&= (\mathbf{y}' - (\mathbf{X}\mathbf{b})')(\mathbf{y} - \mathbf{X}\mathbf{b}) && \text{distribute the transpose} \\&= (\mathbf{y} - \mathbf{b}'\mathbf{X})(\mathbf{y} - \mathbf{X}\mathbf{b}) && \text{substitute } \mathbf{b}'\mathbf{X}' \text{ for } (\mathbf{X}\mathbf{b})' \\&= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} && \text{multiply out} \\&= \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} && \text{simplify}\end{aligned}$$

- To minimize need to calculate derivative of $e'e$ with respect to \mathbf{b} .
- Need to know two things
 - derivative of scalar with respect to vector ($2\mathbf{b}'\mathbf{X}'\mathbf{y}$)
 - derivative of quadratic form ($\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$)

What is the derivative of scalar with respect to vector

- Need to take derivative of $\mathbf{e}'\mathbf{e}$ with respect to \mathbf{b} to find \mathbf{b} that min the sum of squared.
- A derivative of a scalar with respect to a vector

$$y = \mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

$$\frac{\partial y}{\partial \mathbf{x}} = [a_1 \quad a_2 \quad \cdots \quad a_n]'$$

$$\frac{\partial y}{\partial \mathbf{x}} = \mathbf{a}$$

Derivative of a quadratic form

- Equivalent to x^2 is inner product $\mathbf{x}'\mathbf{x}$
- Vector analogue of ax^2 is $\mathbf{x}'\mathbf{A}\mathbf{x}$, where \mathbf{A} is $n \times n$ matrix

$$\frac{\partial ax^2}{\partial x} = 2ax$$
$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

OLS in Matrix Form

Minimizing the objective function

1. Take partial derivative of $S(\mathbf{b})$:

$$\begin{aligned}\frac{\partial S(\mathbf{b})}{\partial \mathbf{b}} &= \frac{\partial}{\partial \mathbf{b}}(\mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}) \\ &= 0 - (2\mathbf{y}'\mathbf{X}) + 2(\mathbf{X}'\mathbf{X})\mathbf{b}\end{aligned}$$

2. Set to 0, and solve for \mathbf{b} :

$$\begin{aligned}\mathbf{X}'\mathbf{X}\mathbf{b} &= \mathbf{X}'\mathbf{y} \\ \mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\end{aligned}$$

What $(X'X)^{-1}$ implies

- For \mathbf{b} to be defined $(X'X)^{-1}$ needs to exist
- $X'X$ must be full rank
- rank of $X'X$ is the same as the rank of X
- The rank of X is between n and $k + 1$, means that $n \geq k + 1$ (obs > variables)
- $k + 1$ columns of X must be linearly independent?
 - Can you have a full set of dummies?
 - Can you include a variable that is always equal to 3?

Takeaways

- Linear regression is the A, B_1, \dots, B_k that solve $\arg \min_{A, B_1, \dots, B_k} \sum E_i^2$
- Solving for linear regression coefficients is relatively **easy**; linear equations; there's an explicit solution. No iteration required.

Linear Regression and CEF

Linear Regression and Normal Distribution

Interpretation

CEF justification for linear regression justification

- Conditional Expectation Function is $E(Y_i | X_i = x)$ for all x
- The CEF is the Min Mean Squared Error (MMSE) predictor of Y_i given X_i
- If the population CEF is linear, then the least squares population regression is the CEF
- If the population CEF is not linear, then the least squares line is the MMSE linear estimate of the CEF.
- See Angrist and Pischke, Ch 3.1

Linear Regression and CEF

Linear Regression and Normal Distribution

Interpretation

But I thought linear regression had to do with the normal distribution?

- Linear regression often presented as

$$y_i = X_i\beta + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma^2)$$

- Why? We haven't had to assume normal distributions before now.
- Helps with statistical inference results.
- However, the CLT handles asymptotic sampling distribution of parameters

Linear Regression and CEF

Linear Regression and Normal Distribution

Interpretation

Interpreting Regression Coefficients β

How the average outcome variable differs, on average:

predictive between **groups of units** that differ by 1 in the relevant explanatory variable while being identical in all other explanatory variables the same

counterfactual in the **same individual** when changing the relevant explanatory variable 1 unit while holding all other explanatory variables the same

See Gelman and Hill, p. 34; Fox, p. 81

References

- Some slides derived from Christopher Adolph *Linear Regression in Matrix Form / Properties & Assumptions of Linear Regression*. Used with permission.
- Material included from
 - Fox Ch 2, 5, 9.1–9.2
 - Angrist and Pischke, Chapter 3.1
 - Gelman and Hil, Chapter 2