POLS/CS&SS 503:
Advanced Quantitative Political Methodology

MATRIX ALGEBRA, LINEAR REGRESSIONS

April 7, 2015

Jeffrey B. Arnold

UNIVERSITY of WASHINGTON
DEPARTMENT OF POLITICAL SCIENCE

CENTER for STATISTICS and the SOCIAL SCIENCES

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 $X_1, \ldots, X_k$.

We will focus on **ordinary least squares (OLS) linear regression**

OLS Objective Function

One X

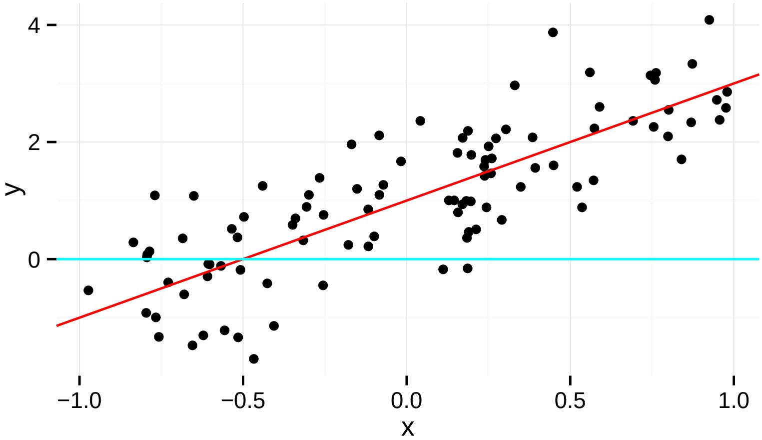
$$A, B = \operatorname*{arg\,min}_{A,B} S(A, B)$$

where

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

How do we minimize this?

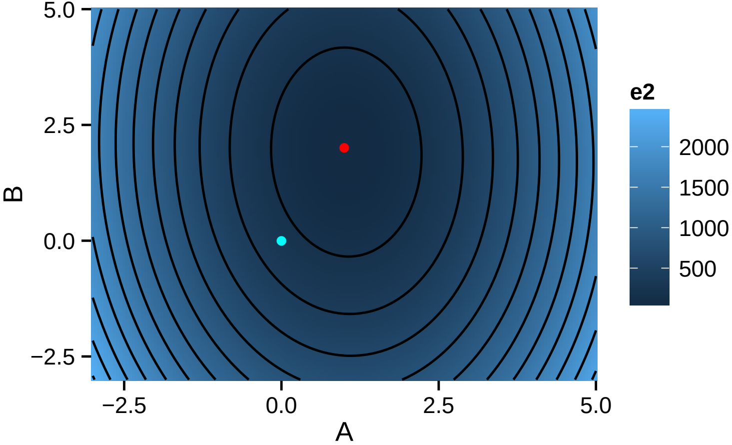
What does the objective function look like?



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 Linear Regression

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

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

Implications of Linear Regression Solution

Least squares A and B

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

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

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

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

Linear Regression Objective Function

Multiple X

When there are k explanatory variables:

$$A, B = \underset{A,B}{\text{arg min}}\, S(A, B_1, \ldots, B_k)$$

where

$$\begin{aligned} S(A, B) &= \sum_i E_i^2 = \sum_i (Y_i - \hat{Y}_i) \\ &= \sum_i (Y_i - A - B_1 X_{i,1} - B_2 X_{i,2} + \cdots + B_k X_{i,k})^2 \end{aligned}$$

How do we minimize this?

Finding the best A, B in Linear Regression

Multiple X

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

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

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

$$\vdots = \vdots$$

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

Not as easy ...

Linear Regression in Matrix Form

Scalar representation

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

Equivalent matrix representation

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

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

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

Want to find the b that minimizes the squared errors:

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

where

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

Why does $\boldsymbol{e}$ 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

$$(\boldsymbol{XB})' = \boldsymbol{B}'\boldsymbol{X}'$$

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

$$? =?$$

Simplify $\boldsymbol{e}'\boldsymbol{c}$

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

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

What is the derivative of scalar with respect to vector

Need to take derivative of $\boldsymbol{e}'\boldsymbol{e}$ with respect to $\boldsymbol{b}$ to find $\boldsymbol{b}$ that min the sum of squared.

A derivative of a scalar with respect to a vector

$$y = \boldsymbol{a}'\boldsymbol{x} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}'$$

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

Derivative of a quadratic form

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

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

Linear Regression in Matrix Form

Minimizing the objective function

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

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

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

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

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

- For $\boldsymbol{b}$ 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 indepdendent?
  - Can you have a full set of dummies?
  - Can you include a variable that is always equal to 3?

Takeaways

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

CEF justification for linear regression justification

- Conditional Expectation Function is $\text{E}(Y_i|X_i = x)$ 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.

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

- Linear regression often presented as

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

- Why? We haven't had to assume normal distributions before now.
- Helps with statistical inference results

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 chaning 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 / Propoerties & 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